EFFECTS OF BEHAVIOURAL INTERVENTIONS ON WRITTEN COMPOSITION

MARIETTE WILHELMINA HOPMAN

A thesis
submitted in partial fulfilment
of the requirements for the Degree
of
MASTER OF ARTS IN PSYCHOLOGY
University of Canterbury
February 1986

CONTENTS

ACKNOWLEDGEMENTS

> "A book - like any piece of writing - is conceived, it gestates, and it's delivery is frequently overdue, accompanied by severe labor pains and followed by postpartum depression. Close relations often suffer sympathetically"
>
> Smith (1982, p.V).

The above process was considerably relieved by the following people, in particular my supervisor Dr. Nirbhay Singh. Sincere thanks Nirb for your generous support and stimulating discussion.

Grateful thanks to Margaret who assisted me in numerous ways, but especially with her humour and friendship.

My thanks to Mrs Fuller and Martin, Mark, Kevin and Brendon of Xavier College; and Mr Kevin Laugesen and Philip, Richard, David and Bevan of Hogben School for their cooperation and hard work.

Thanks also to David Hughes, senior lecturer in Education, for offering his stage two class to holistically mark essays and for useful discussions; to Lianne for assistance with reliability checks; to Dr. Garth Ritchie for assistance with computing; to Tessa Copland for many helpful comments on the manuscript; and to Karilyn Smith for her efficient and reliable typing.

The Arts Library Staff were very helpful, especially the Reference Desk Staff who endlessly ordered my interloans and took an interest.

My final thanks are to a very special man, Tom, who fed, humoured and loved me beyond times of need, and who assisted me so greatly by patiently listening.

ABSTRACT

The purpose of this experiment was to examine the effects of behavioural intervention on written composition. In Experiment 1 correspondence training was applied to the target behaviour of the written output of slow learners. Transfer and maintenance issues were also assessed. In Experiment 2 varying levels of written and verbal content feedback and social reinforcement within a responsive learning context was made contingent on the creative responses of mildly retarded students. In addition the effects of type of picture prompt (action versus theme) was assessed. Both experiments were unique in additionally assessing a comprehensive range of collateral behaviours, through analytic and holistic scoring. Experiment 1 used a changing-criterion design and Experiment 2 an alternating treatments design. The results of Experiment 1 provided strong evidence that correspondence training can effectively be implemented by the classroom teacher, to significantly improve and maintain the written output of compositions. It also led to improved or maintained levels of responding in most collateral behaviours. In Experiment 2 the intervention was effective not only in significantly improving the content and creative ideas, but also in generalizing improvement or maintaining levels of responding in most collateral behaviours. Action picture prompts resulted in significantly more action verbs than theme picture prompts. But in contradiction to literature findings essays with higher levels of action verbs did not always receive highest quality or creative

ratings. Type if picture prompt did not differentially affect collateral behaviours. Analytic scoring was more reliable and offered a more detailed analysis than holistic scoring, but was time inefficient. Attention to the effects of interventions on collateral behaviours and further research areas were identified.

GENERAL INTRODUCTION

What is written composition?

Expression in written form is an abstract mental, social, physical and psychological process which is considered one of the highest forms of language and communication as well as one of the most complex of human acts.

Arnold (1964) defined written composition as an arrangement in writing of that which is experienced, read or imagined and is primarily intended for communication. Applebee (1981) questioned the view that writing was difficult and proposed that the social context was the real difficulty. It was only in certain specialized contexts that writing became difficult, as when they were asked to write for their teachers. Not writing itself, but writing to meet the demands of a particular task complicated the process.

This study has adopted the term written composition. Similar terms, for example writing behaviour, written language, written expression and creative or personal writing reflect the many skills involved in writing, and hence the many definitions needed to cover these skills. Hammill and Poplin (1982) viewed composition as the ability to generate ideas and express them in an acceptable grammar, while adhering to certain stylistic conventions. Walmsley (1983) stressed the communicative function of composition, calling it the intellectual and emotional act of constructing a message. Cramer (1975) defined creative writing as the act of recording ideas in words and sentences. According to Cramer (1975), creativity referred only to the words and sentences as the personal product of the child's experience and

imagination - creative writing was simply an act of personal authorship. Dell (1964) was more specific, saying that creative writing was the expression of students' own feelings and thoughts; it was emotional and sensual.

Creativity has seldom been defined precisely in previous studies. Its definition is a major theoretical issue in this study, because the concept forms the basis for interventions that attempt to manipulate it. This issue will be examined later in the introduction to Experiment 2.

The nature of written composition

The complexity of written composition is evident when one examines the multidimensional skills requiring mastery by the writer. These include conceptual, eye-hand, linguistic and organizational components. Psycholinguists and cognitive psychologists have demonstrated that writing demands the analysis and synthesis of many levels of thinking (Graves, 1978a).

Writing was conceptualized as consisting of transcriptional skills, such as spelling, handwriting, punctuation and grammar, up until the end of the nineteenth century. The concern with mechanical and syntactical features changed however in the 1930s, when composing skills were included in the conceptualization of writing. According to Walmsley (1984) the current concept of composition derives from a rhetorical, a romantic and a cognitive tradition.

The nature of writing and its subsequent evaluation has been perceived in this study as consisting of five interrelated components, including both composing and transcribing skills. This conception is based on that proposed in Hammill and

Larsen's (1983) Test of Written Language. The elements of written composition are:

(a) The mechanical component referring to the skills of handwriting or penmanship;

(b) The productive component, or the quantity of meaningful units in a written passage; for a given message to adequately convey the thoughts and feelings of the writer, an optimal number of sentences must be generated;

(c) The conventional component, defined as the use of accepted rules for punctuation, capitalization and spelling;

(d) The linguistic component, or the use of serviceable syntax and semantic structures; this includes selection of suitable words, tenses, plurals, subject/verb correspondence, and cases; and

(e) The creative component (Hammill & Larsen's 1983 cognitive component) referring to the compositional or imaginative elements of writing. The creative component has often been omitted from studies because of its vagueness, subjectivity and problematic objective assessment.

The skills and abilities required of writing touch all the areas which comprise the language arts, including listening, speaking, and reading. Research has found a close developmental relationship between reading, writing, spelling, speaking and listening (Artley, 1950; Britton, 1972; Dagenais & Beadle, 1984; Emig, 1977; Groff, 1978; Loban, 1976; Lundsteen, 1976; Myklebust, 1965, 1973; Pinsent, 1984). The exception was Fillion, Smith and Swain (1976) who considered these language skills to be independent.

Despite the traditional split in the investigation of reading and writing, the current focus has been on the interrelationship of the two, with the assumption that similar skills underlie both. It has recently been claimed that writing is both facilitative of and enhanced by reading (Bryant & Bradley, 1983; Chomsky, 1971; Graham, 1982; Graves, 1978a, Jacobs, 1984; Lazdowski, 1976; Lickteig, 1981; Martlew, 1983; Murray, 1984; National Council of Teachers of English, 1983; Shanahan, 1980; Smith, 1981; Wiseman, 1984). Literature (Chomsky, 1971; Clay, 1975; Emig, 1977; Graves, 1978a) supporting this assertion has challenged the assumption of early theorists such as Myklebust (1965, 1973) that writing was the last language skill in the hierarchy to be acquired, being dependent on the development of aural, oral and reading language arts (see Barenbaum, 1983 and Walmsley, 1984, for reviews on this debate). Bracewell, Frederiksen and Frederiksen (1982) warned that objective demonstrations that development of composing skills enhances comprehension and vice versa have yet to be achieved. They suggested there was a lack of understanding of precisely what processes the two had in common. The interrelationship between reading and writing is further examined in the symposium in the Language Arts 60(5) May 1983, 545-681; Chall and Jacobs (1983); Clay (1980); Galda (1984); Read (1981); Shanahan (1984); and Walmsley (1984).

Just as writing and reading are related, so too are oral and written language. According to Barenbaum (1983) there has been more emphasis in recent years on the relationship between writing/reading and writing/speaking. Although the nature and function of oral language in relation to

writing is still unknown (Groff, 1978; Reid & Hresko, 1980) the interaction between the two has been established (Barritt & Kroll 1978; Brause, 1979; Cioffi, 1984; Cooper, 1975b; Dagenais & Beadle, 1984; Dyson, 1981; Edmaiston & Larsen, 1983; Green & Morgan, 1981; Hirsch 1977; Loban, 1963, 1976; Smith, 1975; Smith, 1982; Strickland, 1960; Walmsley, 1983). This issue and the debate as to whether oral language training leads to improvements in writing (Emig, 1971; Graves, 1978a) has been reviewed by Emig (1977) and Graham (1982).

What is the role of written composition?

A consistent theme throughout the literature over the past two and a half decades is that composition plays an important and beneficial role in a student's educational and personal success (Funk, 1975). Various writers have given reasons as to why composition is a valid focus in education and research. The reasons included:

(a) Writing being a tool for increasing students' grasp of course material (Pearce, 1983);

(b) Writing as a means of self-expression, a form of communication, a medium for skill development (Jerram, 1985);

(c) Writing helping students to clarify their thoughts, to organize their ideas, and to make sense of the world (Douglass, 1984).

(d) Writing as a means by which students demonstrate their knowledge, and teachers evaluate performance; and as a tool for exploring thought and recording ideas, it

can fulfil emotional needs, and it offers a source of entertainment and enjoyment (Graham, 1982); and finally

(e) Kinnick (1960) stated that "creative writing not only builds better students; it develops better people because it brings to each writer a sense of pride and fulfilment ... [it] sharpens perceptions, expands vocabularies, enlarges concepts acquired through the study of literature, and fulfils the individual student's desire to create ..." (p.22).

According to Tompkins (1982) it has always been assumed that students benefit from creative or story writing, but specific reasons have not been pinpointed. He asked language arts educators, with a special interest in children's writings, why children should be encouraged to write. The seven important benefits suggested were: to entertain; to foster artistic expression; to explore the functions and values of writing; to stimulate imagination; to clarify thinking; to search for identity; and to learn to read and write. Graves (1978a) suggested that writing not only contributed to intelligence and to the development of initiative and courage, but also to reading comprehension and mathematics.

Why the neglect of written composition?

The importance of writing has not been reflected in the quality and quantity of research produced on it. The first major critique and summary of the literature was in 1929, when Lyman argued that composition quality was seemingly so complex as to defy analysis into constituent parts. The research Lyman (1929) reviewed measured pupil products and assumed that by so doing they were evaluating the manifold

processes by which those products were attained (cited in Braddock, 1969). Braddock, Lloyd-Jones and Schoer (1963) compared research in written composition to "chemical research as it emerged from the period of alchemy: some terms are being defined usefully, a number of procedures are being refined, but the field as a whole is laced with dreams, prejudices and makeshift operations" (p.5). Two decades later the situation was little changed. Although there had been a swing to process research, there was still a lack of research-based information on the writing process and instruction. Regarding the written product itself, the research lacked comprehensiveness and ingenuity. Raymond (1982) has suggested that Braddock et al's (1963) image "continued to haunt", and the necessary evolution suggested by the image had not occurred.

Graves (1981) suggested that attention should focus on process, or what occurs when the child writes. He posed ten questions for researchers to investigate in the eighties, and identified the background knowledge necessary to comprehend the writing process. He suggested emphasis be placed on the teacher's role and the need to draw from linguistics, anthropology and developmental psychology. Currently, sources of research on writing are as diverse as psychology, computer science, English and education. Unfortunately Grave's (1981) process-only approach has overlooked the knowledge still to be gained in the other areas of writing. His narrow approach is reflected by a statement that concern with the correct stimulus for writing, with appropriate correcting and grading of final products, and with exercises to increase sentence complexity needs to be abandoned.

One reason researchers have neglected composition is that they have concentrated on reading. Harris (1977) has said that of the three segments of English curriculum (language, literature and composition) the stepchild is composition. According to Walmsley (1984) until 1980, teaching literacy meant teaching children to read, with an absence of writing instruction in any of the basic literacy programmes. Because there was no consensus as to what constituted good writing, unlike reading, what needed to be taught was unclear. Another problem was the assumption that composition simply flowed from the writer, unlike reading which needed to be taught (Brigham, Graubard & Stans, 1972). Graves (1978a) has attributed the disproportionate emphasis on reading in America to political, economic and social forces.

The history of composition research, unlike that of studies in reading, is short, having begun much later. According to Lundsteen (1976) research reports in composition prior to the 1920s were limited in number and scant in their coverage. Over the past quarter century in America, only 156 studies have been conducted in the elementary grades. Of the studies conducted between 1955-1972, 84% were done by dissertation alone (Graves, 1980). Graves (1978b) and Shanahan (1980) have given numerous examples of the neglect of writing using indicators of time, material expenditure, personnel, testing, programmes, articles and writing courses. Researchers (Brigham et al., 1972; Lerner, 1981; Moran 1981) have suggested the neglect is compounded in learning disabled populations who have poor reading and writing skills to begin with. The problems become more severe as the learning disabled child gets older (Dagenais & Beadle, 1984; Poplin,

Gray, Larsen, Banikowski & Mehring, 1980). However Walmsley (1984) has claimed that there is a massive effort to re-balance the writing:reading ratio deficit. Others (Applebee, 1979; Graves, 1980; Hairston, 1982) have also suggested some positive improvements.

The research which has been undertaken in written composition has focussed on mechanical components at the expense of other components (Graves, 1978a).

The lack of research into composition has resulted in many misconceptions arising. O'Dea (1965) discussed five myths in the teaching of composition. His fifth myth, that "students learn to write better by taking into account extensive teacher criticism" (p.330) will be explored in Experiment 2. O'Dea's myths and more still abound 20 years later as Smith (1981) found. His twenty-one misconceptions, or "Smith's myths", revolved around the nature of writing, about how writing is learned, about the act of writing and finally the grand myth about who can teach writing.

Issues in written composition

Steinberg (1980) has described writing as "a garden of opportunities and a thicket of dangers" (p.155). This aptly describes the many issues implicit in the study of written composition (Florio-Ruane, 1983; Stein, 1983). The dominant issues have continued to be teaching methodologies, evaluation and development of a conceptual model and theoretical framework in which to conduct research in written composition. An adequate theoretical account of the link between the processes involved in writing and the resultant text has yet to be offered. Nevertheless many researchers have produced

models to account for the writing process [Barritt & Kroll, 1978; Beach & Bridwell, 1984; Bertram, Collins, Rubin & Gentner, 1982; Britton, 1978; Brown, 1983; Clay, 1975; Cooper, 1981; Cooper & Matsuhashi, 1983; Dagenais & Beadle, 1984; Emig, 1971, 1978; Flower & Hayes, 1981; Frederiksen & Dominic, 1981; Graves, 1975, 1978a, 1983; Humes, 1983; King, 1978; Martlew, 1983; Odell, Cooper & Courts, 1978; Polloway, Patton & Cohen, 1981; Sommers, 1979; Stallard, 1976; Witte, 1980; Young, 1981; Elementary School Journal, 1983, 84(1)]. This latter issue is related to the debate on whether process or product plays a more important role in composition research today.

The emphasis on teaching, measuring and evaluating the finished written product has shifted over the past decade to the writing process (how the product is created). In contrast, in 1963 only two of the 504 composition studies dealt remotely with the process of composition (Brozick, 1978). The term process research refers to "the procedures of prewriting, writing and revising, as well as to denote the psychological, linguistic and cognitive processes employed in the act of composing" (Brozick, 1978, p.84). This research holds the tenets that writing has several distinct stages (e.g., prewriting, writing and editing), that errors are a sign of growth rather than mistakes to be banished, and that the writing sample should be scored holistically or using the primary trait method (Applebee, 1979).

However some process researchers have been highly critical of the product approach. They have attributed assumptions to it that may not in fact exist. For example, Brown (1983) suggested that the experimental methodology

dominating the product approach, emphasizing that which is quantifiable, was "primarily an attempt by researchers in composition to give our field the appearance of hard science" (p.51). She, like Barritt and Kroll (1978), overlooked the fact that a lot of current process research is highly subjective and impressionistic. It relies on case study and interview techniques and statements which lack empirical back-up.

How useful is the process approach? Florio-Ruane (1983) questioned whether enumeration of the processes in a theoretical model of one kind of writing would provide a blueprint for teaching writing to beginners. Further to this, she questioned whether practising discrete skills, without attention to the social context of writing (as models tend to conceptualize the composing process as private) would be of any benefit.

The gradual shift of emphasis in research from the product to the process approach is equally unsatisfactory. Written composition consists of more than these put together, including both evaluation and pedagogical components. An example of the preoccupation with the process model to the exclusion of the product model is evident in a statement by Petty (1978) calling for researchers to "stop examining products as if the processes of composing will leap out at them from the pages they are studying. The product is not behaviour, nor does it represent what has gone on in the individual's mind. It is only a product; process is what people do" (p.83). However de Beaugrande (1984) has dispelled this myth with his work on text (product) analysis which has led to insights into cognitive processes of written composition.

Bereiter and Scardamalia (1983) also acknowledged a direct link between the composing process and written product.

Should one wait for 50 years for research to discover the process, before moving on to develop instruction and evaluation, based on its findings? Teachers need pedagogical research-based strategies now. Issues such as writing conditions (e.g., classroom atmosphere, feedback structures, teacher-child conferences) and various instructional procedural effects on writing need to be addressed. It is unfortunate that researchers and theorists cannot take a more eclectic approach. Process and product are not separate entities; they are interrelated and their separation is detrimental to an overview of written composition.

The preparadigmatic stage of written composition

In Kuhnian terms, writing research is said to be in the preparadigmatic stage, whereby no single approach predominates (Emig, 1978; Gage, 1964; Hairston, 1982; Kuhn, 1970; Young, 1978). Bereiter and Scardamalia (1983) stated that there was no consensus about what needed explaining, what kind of inquiry would lead to explanation, or how it could be judged whether something was or was not an explanation. Evidence of the conflict at the preparadigmatic stage can be seen in the response of critics (Sarbin, 1969) to Zoellner's (1969a) application of behaviour theory to composition. Responding to his critics, Zoellner (1969b) suggested that his article and the response it engendered constituted a classic example of paradigm conflict, and of the way proponents of competing paradigms tended, in Kuhn's phrase, to "talk through" each other, their arguments foundering in a

morass of mutual misunderstanding.

Mosenthal (1982) explored the principles to which an optimal paradigm of elementary classroom writing should conform. According to his article, three principles can critically analyze any theory: The paradigm should define competence in terms of the many different types of writing; yield explanations of the writing process; and define writing competence in terms of the interaction of writing's many contexts. Hairston (1982) has listed 12 principle features of what she considers is the emerging new paradigm for teaching writing.

In the absence of a dominant paradigm, the findings of various research are considered relevant and true, according to Kuhn (1970). In this thesis, the behavioural approach to written composition has been adopted.

The application of behavioural theory to written composition

Prior to the early 1960s, behaviour modification was applied to animals and institutionalized populations. Its application to the classroom (Birnbrauer, Wolf, Kidder & Tague, 1965) began with interventions, followed by research in 1965 (O'Leary & O'Leary, 1976). Studies tended to focus on inappropriate behaviour; however the trend now is towards academic behaviours. This shift was in response to criticism of overemphasizing classroom discipline at the expense of more important goals (Winett & Winkler, 1972). Also, research had failed to support the assumption that reinforcing attentive or on task behaviour would lead to improvement in academic performance (Ferritor, Buckholdt, Hamblin & Smith, 1972; Harris & Sherman, 1974; Hundert, Butcher & Henderson, 1976).

Rather the research had found that direct manipulation of academic contingencies not only led to increases in that behaviour, but also reduced off task disruptive behaviour (Ayllon, Layman & Burke, 1972; Ayllon & Roberts, 1974; Ballard & Glynn, 1975; Broughton & Lahey, 1978; Hay, Hay & Nelson, 1977; Kirby & Shields, 1972; Marholin & Steinman, 1977; Merrett & Wheldall, 1978; Scriven & Glynn, 1983; Van Houten, Morrison, Jarvis & McDonald, 1974; Winnett & Roach, 1973).

Academic behaviours chosen tended to be precise, measureable and objective such as spelling, arithmetic and task completion. However, in recent years some attention has been focussed on complex behaviours such as creative writing. The complexity of creative writing, in terms of definition and measurement, may explain why researchers have avoided the field.

The focus has also been constrained to primary school pupils, to the exclusion of secondary schools in particular (Merrett, 1981; O'Leary & O'Leary, 1976). Experimenting in secondary schools is problematic because, unlike at the primary school, the teacher is not with the class all day, does not know the children well and cannot apply his/her own system. There are also the difficulties of identification and selection of relevant performance contingencies and suitable dependent variables (Harrop, 1978). Another neglected area is special education (Cartwright, 1968; Myklebust, 1973; Poplin et al., 1980; Poteet, 1979; Sedlak & Cartwright, 1972). However this is changing with the influence of the deinstitutionalization or normalization movement, which seeks to give special populations the right to the best educational opportunities and other human rights.

Emphasis in written composition research has been on behaviour analysis and modification. This has proven to be an effective and useful way of modifying the classroom behaviour of a wide range of populations with various academic problems (Copeland & Hall, 1976; Harrop, 1978; Kazdin, 1981; Lahey, 1976; Merrett, 1981; Sharpley & Sharpley, 1981). The underlying assumption is that written composition is a learned behaviour (Carroll, 1964; Emig, 1977; Myklebust, 1965). All behaviour follows similar behavioural principles and hence it is assumed that behavioural techniques can successfully be applied to written composition. The approach depends on writing having the following features (Jerram, 1985):

(a) Writing as a behaviour comes under the control of stimulus conditions or setting events.

(b) Writing is a behaviour which has an effect on the reader and so in time may be influenced or changed by reader response.

(c) Writing behaviour produces consequences for the writer.

(d) Types of feedback as a consequence of writing behaviour.

(e) Writing as the natural reinforcer for writing behaviour.

(f) The output of writing behaviour may be measured objectively, and qualitatively.

Despite some limitations (Kazdin, 1981), behaviour modification procedures have been accepted in the classroom. The procedures have been more readily understood and applied by teachers, than other approaches, because of their demonstrated effectiveness via empirical data; their relevance to academic, management and practical problems; their focus on learning and motivational problems and hence their relevance to educators; and their relevance theoretically to environments

such as the classroom (Copeland & Hall, 1976).

The medical model approach (or deficit model, Bronfenbrenner, 1979) to learning difficulties has speculated on inferred etiological deficits within the individual, not the environment. The behavioural approach, in contrast, has actually alleviated the deficient academic behaviours. Treiber and Lahey (1983) have made the important point that process deficits are not denied; rather that direct treatment of these inferred deficits is neither possible nor necessary for academic improvements to occur. Their assertion was based on empirical evidence.

It is on the basis of empirical evidence that this study has adopted a behavioural approach to written composition.

ANALYSIS OF WRITTEN COMPOSITION

THE ASSESSMENT OF WRITTEN COMPOSITION

In 1963, Braddock et al., commented on the challenge presented by the analysis of composition. This was a common theme even twenty years later. The literature has been consistent in portraying the problematic nature of evaluating written composition (De Shields, Hsieh & Frost, 1984; Gere, 1980; Greenhalgh & Townsend, 1981; Jerabek & Dieterich, 1975; Johnson, 1962; Odell & Cooper, 1980; Stein, 1983). Researchers were faced with reliability and validity problems; they had to decide on and define aspects of written composition to be measured, and then devise techniques to measure them (Cartwright, 1968). Behaviour analysts in particular have had considerable difficulty with the latter two problems. Wotherspoon (1974) stated that such problems have contributed to researchers' hesitancy to take up this field. Despite the advantage of having a tangible product, there was the difficulty of composition being a personal and complex skill. Although the theoretical base of evaluation was unclear, researchers still proceeded to devise ways of evaluating written work.

Composition also brings with it concerns unique to the area of written language. The first is the decision as to how to evaluate the product: whether objectivity should be sacrificed for subjectivity. Secondly, the type of assessment influences the degree to which measurement problems might arise, involving variables such as scorer training, criteria definition and time-consumption. The third factor is whether individual or group design is chosen. Both designs lead to

generalizable information, but individual designs accommodate the writer's unique history and individual differences.

Assessment procedures tend to fall into two broad categories - indirect and direct assessment. Moss, Cole and Khampalikit (1982) perceived the dichotomous distinction as an oversimplification and instead classified written language assessment in terms of task structure, scoring method and level of evaluation.

For indirect, objective or quantitative assessment the subject merely responds to machine scorable items in standardized, multiple-choice and completion tests. These tests are assumed to measure skills indicative of good writing. They are attractive in terms of their cheapness, quick, easy, reliable scoring and good predictive validity (Charney, 1984; Coffman, 1971; Cooper, 1975a). However they lack content and construct validity (Braddock et al., 1963; Cooper & Odell, 1977). They provide an inadequate view of true writing performance, as the behaviour measured does not involve a sample of written composition (Freedman, 1982; Gere, 1980; Odell, 1981; Perkins, 1983). Nor do they test higher skills in writing such as the ability to generate ideas or even larger elements of composition (Braddock et al., 1963). They lack face validity and credibility among English teachers (Breland & Gaynor, 1979; McColly, 1970; Odell, 1981; Perkins, 1983). People are ranked on a scale, but no absolute information about their abilities is provided. The tests involve a passive, reactive mental state, unlike writing which is active. They are incomplete, and lead to a conclusion that form, not content, is more important to teach and learn. Brown (1981, p.37) believes they would make writing courses "cultural programming laboratories".

Applebee (1979) has argued that such usage exercises result in improvements to mechanics, but deterioration in writing, since little practice is obtained in higher level writing skills. In criticizing standardized tests of writing, Cooper (1975a) stated that they don't allow for the delicate, complex human responses to quality, ideas, coherence or style in a piece of writing. Critiques of standardized tests are provided by Dinan (1978) and Stiggins and Bridgeford (1983).

Direct or qualitative assessment of written composition requires that a composition is personally responded to. Direct assessment is more complex than indirect measurement because of the wide range of scoring methods. These methods are said to apply more valid criteria since they assess high level writing skills. However they lack reliability because of the scorer's application of varying standards (Godshalk, Swineford & Coffman, 1966; McColly, 1970). Direct assessment utilises both objective and subjective scoring approaches. The objective approach consists of analytic or atomistic evaluation and computer evaluation. The subjective approach includes various holistic procedures.

Comparison of indirect/direct assessment

With regard to correlations between indirect and direct assessment, Breland and Gaynor (1979) obtained a correlation of 0.63 with college freshmen (N = 819); Godshalk et al. (1966) obtained a range of 0.46-0.75 with high school students (N = 646); Hogan and Mishler (1980) 0.68 (third graders; n = 140) and 0.65 (eighth graders; n = 160) and Moss et al. (1982) obtained ranges of 0.20-0.68 (fourth graders; n = 84);

0.60-0.67 (seventh graders; n = 45) and 0.75-0.76 (tenth graders; n = 98). Although some would conclude from the moderate correlations that both methods tap similar skills (e.g., Breland & Gaynor, 1979), Quellmalz, Capell and Chou (1982) warned that observed correlations were strongly influenced by measure reliabilities. The problem is whether different measures focus on the same text features of compositions and whether they reflect the same underlying skill constructs. The construct validity issue has not been addressed in many comparison studies.

Stiggins (1982) offered a systematic and detailed conceptual analysis and comparison of both direct and indirect procedures based on seven criteria. His review suggested that the major advantages of indirect assessment are high score reliability, low test scoring costs and good control over the skills tested. Advantages of direct evaluation procedures include the information obtained about a subject's proficiency, high fidelity of stimulus and response, adaptability to various writing situations, good face validity and low test development costs. Disadvantages of the direct method are high scoring costs and lack of uniformity of proficiencies. For the direct method, there is a reliance on a subject's reading rather than writing skills, lack of face validity and fidelity to relevant writing tasks. He concluded that the choice of method should be determined according to the user's needs.

Indirect assessment serves only to tap the student's knowledge of language at any point in time. For the purposes of this study, assessment of overall improvement in various writing skills was required and was only achievable through

direct assessment of writing samples. Taking into consideration research which compared indirect and direct evaluation methods, it was decided to use direct methods in this thesis. These methods (analytic and holistic scoring) shall now be reviewed.

Analytic scoring

Analytic scoring is perhaps the most time-consuming and complex of the methods available. The scorer must analyze or count various features of the writing sample to produce an overall score. Analytic evaluation often involves a rating scale, such as the one constructed by Diederich (1974). His system involves eight features: ideas, organization, wording, flavour, usage, punctuation, spelling and handwriting, which are rated on a five-point scale. Advantages of the procedure are that it provides students with a breakdown of their strengths and weaknesses, and that rating scales are said to be highly reliable, although vulnerable to similar threats as holistic scoring. Disadvantages range from the time and commitment necessary to develop an analytical scale for a particular purpose, to their immoderate standard, often vague and arbitrary choice of categories, and contrariety to the principle that the whole may be more than the sum of the parts. In analytic scoring features are scored in isolation and divorced from their context. Another problem is whether error or other counts reflect the overall merit of the written sample. These problems, together with disregard of purpose, speaker role, and conception of audience make analytic scoring's validity questionable.

Computers can be programmed to grade composition,

greatly reducing teacher effort and time. Some perceive them as a rapid and reliable aid, others as an example of the dehumanization of the student-teacher relationship (Slotnick & Knapp, 1971). But can computers evaluate essays as effectively as teachers can? Page and Paulus (1968) found that essays could be processed as reliably by computers as by trained scorers. Slotnick and Knapp (1971) have provided a discussion of the implications and limitations of essay grading by computer.

Holistic scoring

Cooper and Odell (1977) have argued that there is no mechanical or technical solution to the problems posed in evaluating writing. Since writing is an expressive human activity, they believed a "receptive, sympathetic human response" (pp.xii) was best. This philosophy is shared by many teachers and researchers and provides the basis for holistic scoring. This is a quick, impressionistic method in which a single grade or score is given based on an overall appraisal of the quality of a written sample. The final score is the sum of multiple judgements by different scorers. No single feature is directly assessed and penmanship is totally disregarded. As such the procedure cannot supply information as to the strengths and weaknesses of a particular programme or student. However it can emphasize what is right, rather than wrong with a piece. This is based on an assumption of holistic scoring that error alone should not reflect competency (Myers, 1980). The aim then is to evaluate the whole, rather than identify frequency and type of errors. Evidence of this is provided in Perkins' (1980) study which

found that only error-free objective measures or counts (as found in analytical scales) correlated significantly with holistic evaluations.

The holistic approach consists of two propositions according to Alloway (1980): that a writing sample is greater than any of its parts and that no aspect of writing skill can really be judged independently; the halo effect is always strong. Conlan (1978) added a third proposition which is that teachers can recognize good writing.

Brown (1981) questioned the meaning obtained from the holistic score. He suggested that the criteria of quality are not absolute and that the possibility exists for top-ranked papers to still be badly written. In addition, he suggested that traditional holistic scoring is not only unsatisfactory at establishing proficiency in a concrete sense, but also for evaluation of growth.

(a) Reliability:

The subjective nature of holistic scoring makes reliability more of a problem than with objective methods. Poor reliability could result from bias, fatigue, internal lack of consistency, previous knowledge of the student, and/or shifting standards from one paper to the next. Charney (1984) listed other variables likely to threaten reliability: the number of separate readings of each writing sample, the number of writing samples evaluated per student, the writing topic, the size of the rating scale, the consistency with which the readers are trained, and the conditions under which the papers are read. Brown (1981) has stated that it is inadequate as a measuring tool because it is relativistic and not tied to absolute criteria of quality. Hirsch (1977)

has noted that widespread agreement about the qualities of good writing is a precursor of interscorer agreement. Until this occurs, reliability remains substandard.

Evidence on the reliability of holistic scoring is contradictory. Interscorer reliability has been reported to be as high as 0.90 (Braddock et al., 1963; Brown, 1981; Cooper, 1977; Follman & Anderson, 1967; Moslemi, 1975), but other research has demonstrated variability in scorer assessment (Diederich, French & Carlton, 1961; Remondino, 1959; Stein, 1983). Diederich (1974) found that of a sample of 300 essays, one third received grades ranging from one to nine. Such diversity was attributed to individual scorers reacting to certain features they perceived as important in their judgement (Applebee, 1979; Daiker, Kerek & Morenberg, 1978; Hirsch, 1977; Williams, 1980). It was also attributed to scorers adopting different standards of severity and being inconsistent in applying the standards (Jacobs, Zingraf, Wormuth, Hartfiel & Hughey, 1981).

Ways of reducing subjectivity and thus improving reliability have been suggested by various researchers. Jacobs et al. (1981) recommended that researchers: (a) establish criteria to focus scorers' attention on significant aspects of the composition, (b) set a common standard for judging the quality of the writing, (c) select scorers from the same backgrounds, (d) train scorers until they achieve close agreement in their assessments of the same papers, (e) obtain at least two independent grades for each composition, and (f) monitor the scorers periodically during the evaluation to check their consistency in applying the standards and criteria of evaluation. Similarly Henning (1984)

recommended (a) a behaviour-specific rating schedule, (b) an insistence on scorer competence and expertise, (c) the use of multiple independent scorers, and (d) the elicitation of multiple writing samples to control for the fact that attained writing ability may vary with topic and time of day.

High reliabilities are possible when scorers come from similar academic and experiential backgrounds and are trained with a holistic scoring guide (Coffman, 1971; Cooper, 1977; Follman & Anderson, 1967). However even with these two requirements, constraints of cooperation and time could prevent reliable scores.

(b) Validity.

Cooper (1977) regarded holistic evaluation as a valid normative measure but Dilworth and Reising (1979) questioned this. They raised the point that validity issues become complex when a written sample is involved. While the problem of validity in standardized testing ended with the derivation of the answer key, composition assessment extended to the consciousness of the evaluator. Kaczmarek (1980) reported good concurrent validity and Perkins (1983) claimed that of all the evaluation procedures available for written composition, holistic scoring has the highest construct validity. He defined the construct assessed as "overall attained writing proficiency" (p.652).

In contrast, Charney (1984) stated that validity had been assumed and asserted but never adequately demonstrated. In her view, face validity depends on demonstrating that assessment is based on consistent application of acceptable criteria and predictive validity is established by correlating the holistic method with a valid criterion measure taken some

time in the future. She questioned these two validities on two counts. Firstly conditions necessary to ensure reliability may at the same time reduce validity. This could be the case with both choice of topic and scoring procedure. For example, there is evidence to suggest that high reliability could be derived from agreement on superficial features of writing such as handwriting (McColly, 1970) and word choice (Grobe, 1981; Neilson & Piche, 1981; Nold & Freedman, 1977). Although easy to select, these features are irrelevant to true writing skills and as such threaten the validity of holistic scoring. Another issue, perhaps the most central to validity, is posed by the criteria which represent good writing. Charney (1984) described current criteria as having only ad hoc validity; being acceptable only to the group that formulated them. Without widespread agreement and consistent application of criteria, the face validity of qualitative evaluations is doubtful.

With regard to predictive validity, holistic scores have been shown to correlate with quantitative tests (Culpepper & Ramsdell, 1982; Godshalk et al., 1966; Stiggins, 1982). This may not establish validity, but merely reflect that each test has measured similar skills. In addition, the correlation would suggest either that quantitative tests are more valid than previously supposed or that holistic scores are not as valid as was assumed. More importantly, since quantitative methods have been disputed as invalid, they cannot be chosen as a criterion (that is, a previously validated measure) against which to establish the validity of qualitative procedures. The problem of finding an appropriate

criterion measure for cross validation has been identified as a key validation issue (Quellmalz, 1984).

Comparison of analytic/holistic scoring

The comparison of analytic and holistic scoring is complex as the methods have individual sets of assumption rather than being derived from a common set of assumptions about evaluation per se.

Cast (1939, 1940) undertook a comprehensive review of four different methods to assess which was most efficient in marking English composition. She concluded that the analytic method, though laborious and unpopular, appeared uniformly the best. Analytic and holistic marking were the most reliable of the methods, and she suggested that standardized instructions and scorer training could substantially reduce unreliability.

A more recent and superior comparison (Veal & Hudson, 1983) found that holistic scoring was the most economical of the direct measures tested (holistic, analytic, primary trait and mechanics counts) in terms of time and money. Analytic scoring offered a more detailed analysis than holistic scoring, and both correlated highly. For large scale assessment, holistic scoring is recommended because of its face validity, reliability and economy. If detailed feedback is required, this should be substituted with analytic scoring. Veal and Hudson (1983) concluded that analytic and holistic scoring were the most valuable of direct methods available, although both had some disadvantages.

Freedman (1981) found a high correlation between holistic and analytic scoring and recommended the former since

it was less time consuming and produced much the same information as analytic scoring. Prater and Padia (1983a) also found moderate correlations between analytic and holistic scores (0.75; 0.73; 0.60 for expressive, explanatory and persuasive modes respectively) with highest correlations when the category of "ideas" alone was considered.

Because of varying strengths and weaknesses of individual scoring methods, it was decided in this study not to base evaluation solely on any one system, but to modify and adapt existing procedures to suit the needs of the study. Although partly derived from indirect assessment, adapted from the Test of Written Language (TOWL) - an objective standardized test - the final analytic measures focussed on direct assessment of specific skills or elements of written composition. Because of the weaknesses inherent in objective analytic scoring, which forms the basis of evaluation, it was supplemented with an evaluation of a subsample of total essays using subjective holistic scoring.

THE DEVELOPMENT OF A CONTENT ANALYSIS PACKAGE

In addition to conducting Experiments 1 and 2, a third major purpose of this thesis was to develop a comprehensive content analysis of written composition. That this has never been attempted before in the research literature suggests that the task of assessing writing is still in the early stages of development. The content analysis devised needed to be comprehensive because an important intention in Experiments 1 and 2 was to collect data on collateral behaviours to assess unprogrammed effects of the intervention.

It has been recommended (Kazdin, 1973; Wildman & Wildman, 1975) that future researchers collect data not only on target behaviours, but also on collateral behaviours. Research (e.g., Brigham et al., 1972; McLaughlin & Malaby, 1974) has demonstrated that behavioural interventions not only improve academic behaviours, but generalize these improvements to attitudes. Willems (1974) argued that behaviour is part of a delicate system, and when a single behaviour is changed, there are likely to be other concomitant changes. He stressed that these side effects must be monitored when implementing behavioural interventions, so as to assess whether these effects lead to problems over and above those the intervention is supposed to relieve. He considered the behaviour analyst who failed to consider the broader ecosystem to be irresponsible.

Measures were selected on the basis of their reliability, validity and appropriateness. Information was based on TOWL since it is the best test currently available. Former tests of written language such as The Picture Story Language Test (Myklebust, 1965) have suffered from both reliability and validity problems (Anastasiow, 1972; Freedman, 1982; Hammill & Bartel, 1978). TOWL (Hammill & Larsen, 1983) is an easy to administer, standardized, norm-referenced test, with adequate reliability (internal consistency, test-retest or stability and interscorer, standard error of measurement) and validity (content, criterion and construct) (Coleman, 1983; Deno, Marston & Mirkin, 1982; Essex-Sorlie, 1984; Hammill & Larsen, 1983; McLoughlin & Lewis, 1981; Poplin et al., 1980). TOWL consists of both spontaneous and contrived formats covering four of five classes or components of written

skills. The spontaneous subtests do not involve holistic assessment. Further support for the validity of some of the measures adopted in this study comes from a series of studies (Deno, Mirkin & Marston, 1980; Deno et al., 1982) which provide evidence that simple and direct measures (total number of words; mature words; word length and spelling measures) are valid and efficient indices for continuous evaluation of written composition. Although not as complex as some measures, they are economical and produce useful data for evaluating interventions to improve composition.

Objective and subjective measures were selected to comprehensively cover five components, the nature of which was outlined in the introduction. Within the mechanical component, handwriting ability was assessed, guided by the graded samples of Hammill and Larsen's (1983) cursive handwriting scale. Handwriting was defined as the ability to produce correctly formed letters or graphemes, the smallest elements of written language (McLoughlin & Lewis, 1981). Haring, Lovitt, Eaton and Hansen (1978) stated four major drawbacks of the "general excellence" method used by traditional handwriting scales, such as this thesis adopts. They recommended the alternative "factor approach". Since this thesis dealt with many measures, not specifically or solely handwriting, the factor approach was considered too time consuming and not warranted for the present purpose.

The second component was the productive component, which refers to the "quantity of meaningful units in a written passage" (Hammill & Larsen, 1983, p.3). Although not assessed in TOWL, the authors did suggest ways of measuring this component - number of words and thought units (T-units).

Both were adopted in this study. Five measures were subsumed under this component: total number of words, sentences, T-units, new words and vocabulary diversity score.

Total number of words has been adopted as an indicator of maturity in written composition. Freedman (1982) stated that most studies suggest that length is the best predictor of teachers' evaluations. Length has been shown to contribute more to evaluation than mechanical indices (Grobe, 1981; Nold & Freedman, 1977; Slotnick & Knapp, 1971; Stewart & Grobe, 1979) and is considered a better index of idea development (Brigham et al., 1972; Hillerich, 1971). Rubin and Buium (1974) found total words written correlated highly with abstractiveness or creativity and vocabulary richness. Others (Page, 1968; Slotnick, 1972) have also found a relationship between total words written and written performance. The idea of total words written was first developed and reported by Myklebust (1965); and has been shown to validly differentiate normal and learning disabled subjects (Deno et al., 1982; Myklebust, 1973; Poteet, 1979).

A more sophisticated measure than both number of words, and sentences, is number of T-units. Hunt (1966) defined the "shortest allowable sentence" (p.737) or T-unit (standing for minimal terminable units and nicknamed "thought unit") as "a single main clause (or independent clause) plus whatever other subordinate clauses or nonclauses are attached to, or embedded within, that one main clause" (Hunt, 1977, pp.92-3). A clause consists of a subject and finite verb. It is terminable in that it is "grammatically acceptable to terminate [sic] each one with a capital letter at the beginning and a period or question mark at the end"; and minimal in

that they are the "shortest units into which a piece of discourse can be cut without leaving any sentence fragments as residue" (Hunt, 1966, p.737). The measure adopted in this study is a modification of Hunt's (1965) Terminable Unit, which is simpler to score but provides similar information to his more complex measures. The modified measure can be defined as a "segment of meaningful expression that contains an identifiable verb and it's subject and that can stand alone" (Hammill & Larsen, 1983, p.44). Progress as a result of intervention would be reflected in an increase in word output, sentences and T-units.

Another productive measure is number of new words - that is, words never before used by a student in an essay in this study. Only one other study has adopted this measure (Brigham et al., 1972). In Brigham's study the new-word contingency did not increase new words; rather, they declined.

The final measure within this component, vocabulary diversity or "different words" (Brigham et al., 1972) was included to assess firstly whether the intervention influenced vocabulary richness, and second whether the subjects increased their word production by word repetition (redundancy). Brigham et al. (1972) in fact made reinforcement contingent on the production of different words, so as to decrease within-story redundancy, but this had little effect on the production of different words. It should be noted, however, that Clay (1975) perceives repetition positively. She views such behaviour as an indication of progress and a feature of emergent writing behaviour. Vocabulary diversity was measured in this study using a modification of the statistic called the "type-token ratio" (TTR) in which the ratio of different words

(types) was divided by the total words (tokens) (Johnson, 1944). Because it is sensitive to the total number of words in a sample, some have computed the ratio based on only the first 50 words of each composition, termed TTR:50 (Cartwright, 1968; Sedlak & Cartwright, 1972). However Carroll (1964) suggested his corrected ratio, independent of sample size and based on the total number of words is a more reliable measure. It has been found to be relatively free of bias from sample size fluctuations (Sedlak & Cartwright, 1972) and the corrected ratio is said to increase across grade levels (Ciani, 1976).

The third component, the conventional component, is concerned primarily with the measures of capitalization, punctuation and spelling. Hammill and Larsen (1983) used the term "conventional" because these measures are based on rules which govern their use. The TOWL assesses these measures through contrived tests, but does offer excellent charts (Burns, 1974) against which to compare a spontaneous piece of writing for punctuation and capitalization errors. The only weakness with this approach is that spontaneous writing samples do not necessarily incorporate usage of the full range of capitalization and punctuation rules. However they do indicate how the child applies these rules to natural, day-to-day writing, not found in a contrived test. Rather than follow the common tendency to focus on error, punctuation and capitalization measures have been described in terms of appropriate, or correct usage (Bording, McLaughlin & Williams, 1984).

Spelling has been defined as the "ability to produce the sequences of letters that form words" (McLoughlin & Lewis,

1981, p.440) and is imperative to communication of ideas. The incorporation of this measure made it possible to assess if spelling errors increased as a function of increasing the required number of words to be written (Experiment 1). It gave insight as to whether the subjects were so preoccupied with keeping spelling accuracy at a constant rate that few risks were taken in introducing low frequency or less common words to the discourse (Croft, 1982; Glynn, 1981).

Barr and Lambourne (1984) commented that the true nature and complexity of spelling skill gained through instruction and demonstrated in producing written composition, has been largely ignored, as has the focus of the spelling of a single individual over several writing samples where the writing task is held constant. This neglect is unfortunate, since research has demonstrated that different processes are involved in spelling performed in a spelling lesson or test (the decontextualised situation) and in a spontaneous writing situation, in terms of correlation (Croft, 1982) and errors specific to each format (Hotopf, 1980; Smith, 1983; Sterling & Smith, 1981; Wing & Baddeley, 1980). This may stem from their involvement at different levels of complexity.

The scorer is also influenced by whether a test or written sample is marked. On a spelling test one only has to assess whether or not the word is correctly spelt and if not, the number of errors made. In contrast, many more explicit decisions must be made in the writing task and consistently adhered to. For example "What was the intended word? Is the error a spelling error as opposed to a handwriting or grammatical error? Should the assessment be based on the number of different words misspelled? Can spelling be examined in

isolation from such features as the amount written, the type of words used, the rate of production, the success of communication, the boldness of the style?" (Barr & Lambourne, 1984, p.298).

The linguistic component is one of semantics and syntax. Semantics refers to the intentions, ideas and thoughts being expressed in a written passage, typically considered as the appropriate and meaningful use of vocabulary (Hammill & Larsen, 1983). It includes features such as vocabulary, style, organization, creativity and productivity (McLoughlin & Lewis, 1981). Syntax is concerned with the arrangement and interrelationship of words in sentences (Dixon, 1972) and has been defined as the ability to write sentences and paragraphs that are grammatically correct (McLoughlin & Lewis, 1981). Because syntax and semantics deal with higher-level functions than previous measures, they are more problematic to measure (Hammill & Larson, 1983).

The first semantic measure to be assessed was the vocabulary score. The procedure simply involves counting the total number of seven or more letter words in a sample. Research has suggested that word length is related to skill in written expression (Deno et al., 1982; Page, 1968; Siegelman, 1983) and that long words are both more rarely used than short words and used more frequently by mature readers and writers than younger ones (Carroll, 1964; Fry, 1977). Additional assessment of vocabulary was gained using a modified simple rating system (Burns, 1980) called the "word usage score". The "mature usage score" was also obtained. Mature or infrequently used words did not appear on a list of undistinguished, frequently used or immature word choices. The

list was devised by computer analysis of numerous written compositions of 4th, 8th and 11th graders with topic-related words being taken into account (Finn, 1977).

Syntax was assessed by counting the number of grammar errors per 100 words, with the assistance of an analysis chart of grammar (Burns, 1974). The syntax measures of sentence complexity and immature connecting words (Hammill & Larsen, 1983) were not assessed because of their time-consuming nature and irrelevance for the standards of writing assessed in this thesis.

In the 1960s the most widely used measure of syntactic maturity - sentence length - was replaced by Hunt's (1965) measure of mean T-unit length. Of four indices of syntactic maturity measured, Hunt (1965) found this the most valid, especially in early grades, and mean sentence length the least useful. Sentence length was an unreliable index because it varied with punctuation skill and writing style. Sentences could be long because of confounding independent clauses, not because of structural sophistication. In contrast the T-unit avoided length imposed by coordinate clauses, and so was more reflective of syntactic sophistication (Dixon, 1972). Hunt (1965) found the T-unit more complex with length, because of increased clause length in younger children and increased subordination in older children. According to Gripps and Ewen (1974) mean T-unit length had the advantage of accounting for all the subordination in a written passage and all the coordination between words, phrases and subordinate clauses, although it did not show the coordination between main clauses. Hunt (1970) found that while length varied irregularly, T-unit length was shown

to increase with age, and T-units per sentence to decrease.

Hunt (1965, 1966, 1977) empirically validated claims made about the T-unit concept by testing two assumptions: that mean T-unit length and the number of dependent clauses per T-unit increased with age, as more sentence constituents were consolidated into the T-units (Mosenthal, 1982). Hunt (1965) found an increase in mean T-unit length across grades 4, 8 and 12, and also that shorter T-units were written by high school students than by skilled adults. In a later study (Hunt, 1970), in which subject matter was controlled, T-unit length was also found to increase with age and ability (grades 4, 8, 10, 12). Smith's (1974) study replicated and affirmed Hunt's (1970) findings.

Support for Hunt's findings came from O'Donnell, Griffin and Norris (1967) who described the T-unit as a simple, objective and valid index of syntactic maturity. They established norms for children younger than and overlapping with Hunt's sample (1965, 1970) and found a similar increase in mean T-unit length across grades, as other studies (Braun & Klassen, 1973; Faigley, 1977; Morris & Crump, 1982; Stewart, 1978; Stewart & Grobe, 1979; Veal, 1974; Watson, 1980). Loban (1976) followed kindergarten children for 13 years and demonstrated that the T-unit was a good measure of language sophistication. High mean T-unit lengths were not attributed to an increase in meaningless language, that is verbosity, but rather were associated with high ratings of language skill. A more effective use of phrases and clauses, and increased use of other forms of elaboration contributed to clear and meaningful communication. Others (Deno et al., 1982; Dixon, 1972; Mellon, 1969; O'Hare, 1973; Perron, 1977;

Potter, 1967; Veal, 1974) have included the measure in their evaluation of written composition.

Recent studies have found that task variables affect syntactic complexity in written composition, namely mode of discourse (Crowhurst & Piche, 1979; Johnson, 1967; Perron, 1977; San Jose, 1973, 1978; Veal & Tillman, 1971; Watson, 1980; Witte & Davis, 1980) and audience (Crowhurst & Piche, 1979; Perron, 1979).

The implications of these findings suggest that normative data need to be revised to take into account these task variables. Composition researchers (Hunt, 1965, 1970; Mellon, 1969; O'Donnell et al., 1967; O'Hare, 1973) largely ignored the fact that variables other than growth or maturity, due to writer age, could affect syntactic maturity. This ignorance has been attributed to their assumption that T-unit length was a stable individual trait (Combs, 1978; O'Donnell, 1976; Witte & Davis, 1980).

Other criticisms have been that Hunt's T-unit analyses were sensitive neither to meaning nor mechanical errors (Gregg, 1983) nor to change (Deno et al., 1980), and that they fail to deal with excessive coordination within a sentence (Ney, 1966) and issues related to the validity of the T-unit. O'Donnell (1976) suggested that mean T-unit length did not discriminate among the various ways length could be achieved, some of which required more linguistic maturity than others. Odell (1979) suggested that Hunt's notion of syntactic maturity lacked a clear relationship to the writer's cognitive maturity. Growth in writing was considered but growth in cognitive processes ignored. Moffett (1968) criticized it for basing sentence maturity solely on sentence complexity;

Hirsch (1977) because it neither assessed nor measured the writer's intent to communicate meaning; and Perkins (1983) because, like other objective measures, it failed to quantify cohesion, coherence, organization, idiom, diction, tone, relevance, or focus - all of which contributed to good writing. Finally, Mosenthal (1982) suggested that T-unit analysis and similar descriptive grammars characterized only linguistic performance and overlooked linguistic competence.

Despite the limitations of the index, O'Donnell (1976) concluded that T-unit length was still the most useful and useable index of syntactic development over a wide age range. To assess the impact of an intervention on written syntactic fluency, Cooper (1975a) suggested that an increase of 0.25 to 0.50 words per T-unit per year could be expected. Results from Experiments 1 and 2 will be analyzed to assess if this in fact occurred.

Syntax also involves parts of speech. The four major groups, nouns, action verbs, adjectives and adverbs, have been assessed in this study.

The creative component, concerned with the conveyance of ideas and meaning, was assessed by two measures. The first was a modified version of the thematic maturity measure of Hammill and Larsen (1983). Because this measure had received criticism, for the inclusion of certain indices, as well as for the global all-or-nothing nature of the scoring, it was to be supplemented by the Carlson Analytical Originality Scale (Carlson, 1961, 1963, 1965, 1973). Carlson (1961) defined original writing as a form of writing which was individual, novel or unusual. Originality could be expressed through the dimensions of story structure, novelty, emotion,

individuality and story style, which her scale covered. A review of the scale was unable to be located.

The sixth and final component to be assessed was the contact component. The sole measure used was the number of words the writer asked how to spell. The measure applied to Experiment 2 only. It was monitored to assess whether contact would increase in proportion to the rate of production, or mature words attempted, for example.

CONTROL OF FOUR MAJOR VARIABLES

Four major variables need to be controlled for when directly assessing writing samples: the writer and assignment variables and intra and interscorer reliability. These issues will be addressed individually. Unlike the majority of researchers who view variability negatively, Graves (1983) viewed it as a positive trait. He stated that high variability reflected that writers were given the opportunity to take risks, choose their subjects, and experiment with ways to say things. He contrasted this to unhealthy variance when children took few, if any, risks because of overattention to mechanics or if they were punished or received minimal commentary.

It is often assumed that threats to reliability reside in the scorer, but the writer can influence measurement stability as well. The writer variable concerns day-to-day variations in the writing performance of an individual. Graves (1983) suggested that it was a myth that all writing samples were of the same quality and this statement was empirically supported mainly by early research. These

researchers consistently found high variability between two written samples of the same writer (Anderson, 1960; Kincaid, 1953; Martin, 1961; Traxler & Anderson, 1935). They concluded that one writing sample was an unreliable and invalid basis for evaluation (Braddock et al., 1963; Diederich, 1974; Freedman, 1982; Henning, 1984; Raymond, 1982). Dilworth and Reising (1979) suggested that variables such as fatigue, interest and personal attitudes contribute to writer variability and Graham (1982) suggested changes in health, and distractions. Graves (1983) listed eight major categories, including self-concept, process and organic base, with the remaining five relevant to the assignment variable. However he cited little evidence for these opinions.

In order to control for the writer variable in Experiment 1 approximately 70 samples were collected from each subject and 32 samples from each subject in Experiment 2.

Braddock et al. (1963) stated that four aspects of the assignment variable needed to be controlled to minimize variations in writing quality: the topic; mode of discourse; time afforded for writing; and the situation, for example the day it was written, lighting, and time of day (Dittmer, 1974; Diederich et al., 1961). Additional factors included the source of stimulation, teacher directions, the intended audience and the significance of the situation (Graham, 1982).

The majority of empirical evidence has focussed on only one of these variables, the selection of writing topic, and suggests that topic affects performance (Freedman, 1981; Freedman & Calfee, 1983). This area is controversal as there is no consensus on whether the topic should be broad or narrowly defined (Charney, 1984). McColly (1970) for example

argued that an open, less structured topic allows for a wider range of responses, and lowers validity and reliability of ratings. He suggested that not only should topics specify the discourse aim, they should also be selected to control for the effects of knowledge. However Lloyd-Jones (1977) warned that restricting the situation in order to define a purpose and stimulate performance of a particular kind would increase the chance that the exercise fell outside of respondents' experiences. Nold (1981) argued that topics with different discourse aims should not be compared.

Myers (1980) suggested that topics be screened for problems of focus, special knowledge, open-endedness and grade-level differences. Both Myer (1980) and Freedman (1981) recommended that topics be pretested to ensure equivalence with previous topics.

With reference to discourse aim, research (Godshalk et al., 1966; Quellmalz et al., 1982; Prater & Padia, 1980, 1983b; Veal & Tillman, 1971) has suggested that the mode of discourse required by a topic leads to skills intrinsic to that mode, and hence variability when tasks contain different discourse aims. Researchers must therefore be aware of the mode of discourse their topics elicit.

Topic selection was also affected by the scoring procedure adopted. Holistic scoring for example allowed for wide-open topics and different discourse modes to be treated similarly. In contrast, other scoring procedures were developed for specific discourse modes, which required specific topics to elicit them (Cooper, 1977; Lloyd-Jones, 1977). Hence the criteria which constituted a valid topic would differ depending on one's scoring procedure.

Because of this, Hagstrum (1964) strongly criticized Kincaid's (1953) study which demonstrated that varying topic or discourse mode had little effect on writing quality. Hagstrum attributed this contradictory finding to flaws in the study, especially the validity of the topics assigned.

The assignment variable was controlled for in Experiments 1 and 2 by providing similar topic material, studying only one discourse mode, that is narrative writing, allowing writing to occur in 30 minutes only, writing on similar days of the week, in the same room with the same lighting and teacher or experimenter, and standardizing directions.

The last two variables pertained to the reliability of scoring the written sample and concerned variations in a scorer's standards of evaluation. Empirical evidence, dating back to the 1880s, suggests that consistency in intra and interscorer reliability is difficult to achieve (Coffman, 1976; Ebel & Damrin, 1960; Follman & Anderson, 1967; Huddleston, 1954; Meckel, 1967; Smith, 1969).

Intrascorer variability refers to the same scorer assigning different grades to the same writing sample on different occasions (Coffman, 1976). This has been attributed to scorer fatigue, personal feelings, personal familiarity with the purpose behind evaluation, and unconscious or conscious response to various factors (Braddock et al., 1963; Graham, 1982).

Although research has demonstrated the occurrence of low intrascorer reliability (Follman & Anderson, 1967; Hartog & Rhodes, 1936; D.C. Hughes, personal communication, 1985), the majority of research has been in the area of interscorer reliability.

Several explanations for low interscorer reliability have been offered. Firstly, individual scorers differ in the severity of marking, tough scorers giving low grades and easy scorers high grades (Freedman, 1981; Myers, McConville & Coffman, 1966). Related to this issue is the fact that scorers also differ in the range of marks they distribute. Some use the whole score scale, others a narrow range and still others score around their average (Coffman, 1976; D.C. Hughes, personal communication, 1985).

A second explanation is the phenomenon of proactive effects, also known as contrast or context effects. The quality of papers or essays preceding subsequent papers, affect the evaluation of those subsequent essays in either a positive or negative direction, depending on the quality of the preceding papers. This assumption was initially demonstrated on essay question responses and results suggested that a block of very good essays at the start of a sample of essays, depresses subsequent essay scores, and a block of poor ones, enhances scores on subsequent essays (Hales & Tokar, 1975). Other studies (Daly & Dickson-Markman, 1982; Hughes, Keeling & Tuck, 1980a) empirically support the phenomenon.

Suggestions have been made on how to reduce the contrast effect, for example randomly shuffling papers before grading them (Daly & Dickson-Markman, 1982; Hales & Tokar, 1975); or reading a sample of essays prior to grading (Hughes, Keeling & Tuck, 1980b). The latter method and analytic scoring were unsuccessful in minimizing context effects (Hughes et al., 1980a). Another procedure, instructing scorers to guard against context effects and sorting and rereading essays

prior to grading, also failed to reduce or eliminate the effects (Hughes, Keeling & Tuck, 1983). A final procedure, providing scorers with model essays so that they could refer to these during scoring, has also proved unsuccessful (Hughes & Keeling, 1984).

A third explanation for low interscorer reliability may be that scorers use different criteria by which to judge the essays. Research has found that teachers do not agree on which elements define good writing (Diederich et al., 1961; Dittmer, 1974). The most important elements to teachers were variously essay length and minimal spelling errors (Stewart & Grobe, 1979); vocabulary characteristics (Grobe, 1981); development and organization of ideas (Freedman, 1979a, 1979b) or number and novelty of sentences in an essay (Malgady & Barcher, 1977). This variability would in turn influence their assessment.

Although in theory teachers stated that context and organization of ideas were most important, in practice if a paper had poor mechanics, despite good ideas, it would be downgraded (Harris, 1977; Martin, 1975; Scannell & Marshall, 1966). Lower scores were also assigned by elementary school teachers (Markham, 1976), secondary (Briggs, 1970; Chase, 1968; Soloff, 1973) and prospective teachers (Marshall & Powers, 1969), when handwriting was of poor quality, irrespective of content quality. Huck and Bounds (1972) investigating this issue further, concluded that essay neatness interacts with the handwriting neatness of the person scoring the essay. Scorers with neat handwriting would downgrade an essay with poor handwriting, while scorers with messy handwriting would not. Handwriting also influenced the assessment

of essays by holistic scoring (Briggs, 1970). Only one study (Marshall, 1972) found no significant differences with varying levels of writing neatness and spelling errors.

Freedman (1979a, 1979b) also found that variables within the essay contributed significantly to the variance of holistic scores. Two environmental variables - the training of scorers and choice of topic - were the next most influential. This view is supported by other researchers (Moslemi, 1975; Raymond, 1982) who concluded that the quality of writing is multidetermined by interactions among the text, authors and readers. Because of this Raymond (1982) has suggested that interscorer reliability of around 0.80 is to be expected and not apologized for and higher rates are to be treated with suspicion.

Having acknowledged the occurrence of multiple threats to reliability, they need to be controlled and minimized. In fact guidelines for reducing scoring error have been suggested (Coffman, 1976; Graham, 1982; Moslemi 1975).

Researchers have consistently agreed that a set of clear predetermined criteria which are understood and endorsed are an essential step in achieving reliability (Braddock et al., 1963; Buxton, 1958; Charney, 1984; Graham, 1982; Jacobs et al., 1981; Meredith & Williams, 1984; Moslemi, 1975; Quellmalz, 1984).

Although research suggests that holistic and analytic scoring are equally reliable, some have claimed that analytic scoring is more reliable (e.g., Cast, 1939, 1940). They state that including a rating scale is an important step in achieving high interscorer reliability (Coffman, 1976; Moslemi, 1975). It has been suggested the rating scale should have fine rather

than gross levels; however there is little consensus on the number of points or intervals (Graham, 1982). The more points a scale has the more imperative it is to develop a clear and defined reference point so as to improve reliability (Coffman, 1976). McColly (1970) has suggested avoiding odd-numbered scales as they tend to shift ratings to a midpoint. Diederich (1974) developed a well-known analytic rating scale to overcome problematic interscorer reliability. However there is no evidence that agreement is better using this scale than with holistic scoring.

Interscorer reliability improves when scorers receive orientation training and practice in applying the standard criteria and make periodic checks on their rating (Bridwell, 1980; Buxton, 1958; Charney, 1984; Cooper, 1977; Diederich, 1974; Ebel & Damrin, 1960; Graham, 1982; Jacobs et al., 1981; Mellon, 1975; Meredith & Williams, 1984; Myers, 1980; Moslemi, 1975; Quellmalz, 1984). Multiple ratings by at least two independent scorers are recommended (Coffman, 1976; Graham, 1982; Henning, 1984; Jacobs et al., 1981) and they should come from similar backgrounds (Bridwell, 1980; Cooper, 1977; Follman & Anderson, 1967; Jacobs et al., 1981; McColly, 1970). But even if scorers are homogeneous, they still require training to achieve agreement (Diederich et al., 1961).

It has been argued that homogeneous scorers (from similar academic backgrounds) shared common experience, attitudes and values. Follman and Anderson (1967) hypothesized that when scorers are homogeneous, their ratings will be consistent, whether or not they used a rating scale. However, when scorers are heterogeneous, it is more imperative to use a rating scale. Without it essays were rated according to

different, idiosyncratic elements. The scale functions as a sensitizer or organizer of the perception, and gives direction to his attitudes and values. It focuses the attention and guides judgement. Follman and Anderson's hypothesis has been empirically supported by two studies. Moslemi (1975) using a rating scale, and Smith (1969) using a STEP Essay Test, found that a homogeneous academic background, experience or professional training, are not prerequisites for achieving inter-scorer consistency.

> Tis with our judgments as our watches: none
> Go just alike, yet each believes his own
> A. Pope, Essay on Criticism.
> (cited in Hirsch, 1977, p.176)

This review has demonstrated that threats to reliability and validity are complex, yet need to be controlled for in the assessment of written composition. This has been achieved in several ways in Experiments 1 and 2. With respect to intra and interscorer reliability, each sample was assessed using consensus grading, whereby two scorers, having undergone a training course provided by this writer, marked essay samples together. This enabled discussion in doubtful and ambiguous situations and control of each other's deviations (Diederich, 1974; Moslemi, 1975). The procedure proved time consuming but excellent for this experimenter's purposes. Twenty-five percent of the written samples were then marked a third time by a single scorer to obtain reliability coefficients. Scorers were monitored periodically during evaluation to check that criteria and rules were being consistently applied (Jacobs et al., 1981; Moslemi, 1975). In the main, analytic scales and mechanical counts were adopted because of their objectivity and clearly defined criteria and rules.

Analytic scoring was supplemented with holistic scoring of a portion of the total essays. The scoring was done by a stage two university education class. The essays in Experiment 1 were graded by four different scorers, and in Experiment 2 by three different scorers. By assessing interscorer reliability, the claims that global, subjective holistic scoring was less reliable than analytic scoring could be tested.

EXPERIMENT 1

INTRODUCTION

Behavioural procedures involving reinforcement contingencies have been applied to various academic areas, namely reading (Singh, Singh & Winton, 1984; Singh, Winton & Singh, 1985), spelling (Foxx & Jones, 1978; Ollendick, Matson, Esveldt-Dawson & Shapiro, 1980), arithmetic (Broughton & Lahey, 1978; Van Houten & Thompson, 1976), social studies (Harris & Sherman, 1974) and history and geography (Glynn, 1970).

Within the academic area of written composition, reinforcement contingencies have successfully been applied to various domains.

Penmanship, or handwriting, has attracted some research interest over the past decade. Reinforcement techniques have focussed on increasing the rate and/or quality of printing and/or cursive writing. Hopkins, Shutte and Garton (1971) focussed on all four areas with first and second graders. Salzberg, Wheeler, Devar and Hopkins (1971) and Rapport and Bostow (1976) found access to play contingent on increased accuracy effective in a kindergarten setting. However Salzberg et al. (1971) found little evidence of letter reinforcing generalization, contrary to Lovitt (1976). Lovitt (1976) obtained transfer of cursive writing skills on all letters of the alphabet after drilling only a quarter of them, as well as transfer across type of writing, that is from cursive to manuscript.

Point systems have also been found to increase handwriting completion rate (McLaughlin & Malaby, 1972, 1974). Trap, Milner-Davis, Joseph and Cooper (1978) applied three treatment conditions to normal first graders: (a) verbal and visual feedback and praise; (b) treatment (a) plus rewriting of incorrect letters; (c) treatment (a) plus a certificate. Although this resulted in improved cursive handwriting, the component(s) responsible were not able to be determined from the design.

The earliest study with special populations focussed on initiating writing skills in mentally retarded children (Bijou, Birnbrauer, Kidder & Tague, 1966; Rayek & Nesselroad, 1972). Miller and Schneider (1970) applied token reinforcement to prerequisite writing skills of Head Start preschoolers. The skills included pencil holding and stroke drawing. Collateral behaviour changes included development of a favourable attitude to school, an increase in cooperative play with peers, and improvement in vocabulary and understanding of instructions. Fauke, Burnett, Powers and Sulzer-Azaroff (1973) applied an instructional and reinforcement package to a 6-year-old boy with behaviour problems and poor printing. This resulted in the subject achieving 100% proficiency in oral letter identification and in writing letters with and without a model. This also generalized from the non-school to classroom setting. Nichols (1970) also found that token reinforcement improved the writing skills of both slow learners and subjects with discipline problems. Flash card modelling and social reinforcement procedures eliminated letter reversal, omission and substitution, and digit reversal errors in learning disabled

children (Stromer, 1977). And finally, Masterson, Earls and Taylor (1981) successfully improved the cursive handwriting of a mildly mentally handicapped teenager.

Thus, applied behaviour analysis, particularly contingent reinforcement, can successfully be applied to the handwriting aspect of written composition. For a more detailed review of such studies and methodological issues, the reader is referred to Hansen (1978) and Kerr and Lambert (1982).

Nichols (1970) reinforced slow learners and discipline-problem pupils for successful completion of three writing tasks: (a) a perfectly neat paper with no missing words; (b) no crossed out words; (c) correct sentence structure and no misspelled words. The usefulness of the tasks, however, was questionable. Errors are a normal occurrence, especially with new tasks and topics (Mellon, 1975; Shaughnessy, 1977). Writing skills and independence increased as assessed objectively and subjectively.

Capitalization and punctuation in adolescent mildly retarded students (Bording et al., 1984) and prepositional phrases (Brigham, Burt & Edwards, 1976) with a third-fourth grade average class have also been researched. In both studies, consequential procedures were combined with antecedent instructions. Rate, accuracy and prenomial adjectives and adverbs all increased when reinforcement was supplemented with modelling and remedial feedback, in hearing-impaired and aphasic 8 to 10-year-olds (Heward & Eachus, 1979). In a telephone-managed home-based summer programme for 13 to 14-year-old learning disabled students, both action words and adjectives increased. A treatment package (consisting of

praise for self-selecting and self-locating words from supplied word lists, and response cost for words shown or told by the teacher) increased child self-selected words and decreased words shown by the teacher (Wilson, 1976).

Fluency in written expression has also been a target behaviour in composition research. A daily independent spelling programme was also conducted, and resulted not only in improvements in the written expression and spelling of the 11-year-old under-achieving male, but also in generalization of the spelling programme into the writing programme (Glynn, 1981).

Regarding contingencies on productivity, Eachus (1971) found that token reinforcement and verbal remediation resulted in higher response rates and levels of accuracy, and an increase in the frequency of higher-order sentence writing in bright, deaf fourth graders. Brigham et al. (1972) and Glendinning (1977) both found that story length increased when sequential token reinforcement and minimal instruction were contingent on number of words written. Other response components were also manipulated and in both studies attitudes and subjective ratings of quality improved.

Van Houten et al. (1974) focussed on the effect of precise academic feedback of explicit timing, immediate self-scored feedback, and public posting of results and instructions on the compositional response rate of second and fifth graders. This was an alternative to the usual intervention of token economies which the authors considered too expensive and time consuming for teacher intervention. The intervention resulted in both a doubling of word rate and an increase in

subjective ratings of quality. However the longterm benefits of asking subjects to write as many words as they could in ten minutes was questionable. As with other researchers, subjective improvement in attitude was observed, but not supported by objective data. Unfortunately the study did not assess which combination of the four intervention components or perhaps indirect effects of the intervention such as attention from teacher and peers (as social interaction increased) was responsible for the results.

The issue of causation was further examined by Van Houten, Hill and Parsons (1975) with regard to composition, reading and language exercises. Timing and feedback increased performance, public posting of results led to an even greater effect, and teacher praise led to even more improvement in one of two fifth grade classes. The class for which praise was ineffective wrote for only ten minutes, unlike the other class for which praise was effective. They wrote for twenty minutes, thus giving them more opportunity to increase rate and therefore more probability of access to teacher praise. Performance comments by the children on their own and their peers' work increased in the intervention phases. Interestingly, only one of the two scorers perceived an increase in story quality. Details of how many stories were rated, description of the scorers, whether or not they received training, and interscorer reliability coefficients were not provided.

Van Houten and McKillop (1977) implemented the performance feedback system of timing, self-scoring and public posting to grade 10 and 11 students' story writing. The number of words increased as a consequence of the

experimental package and these essays were rated superior to baseline stories. In a later study (Van Houten, 1979), the performance feedback package sequentially applied to words per minute, again demonstrated that this intervention could increase word output.

McKessar (1977) found that verbal and written comments and group token economy improved the writing rate, accuracy and quality of a normal form one class. The contingency did not lead to repetitious writing to maximize reinforcement, and a positive change in pupil attitude was noted. The writing session was only five minutes long.

Kraetsch (1981) used oral instructions and praise to increase total number of words, sentences and vocabulary diversity, as well as number of adjectives used by a 12-year-old learning disabled boy. However compound/complex sentences simplified. Scriven and Glynn (1983) assessed the effect of a systematic performance feedback programme on written prose, skills, and items tasks with low-achieving fourth formers. Not only were there major gains in rate, accuracy and task completion across all three tasks, but the mean level of on-task behaviour also increased, variability decreased, and positive changes in teacher behaviour were noted.

Glynn (1981, 1982a) has suggested that contingencies on accuracy, before fluency is achieved, could be counter-productive particularly with low achievers. This could possibly result in writing fluency failing to establish, and even in writing extinguishing. This view is supported by applied behavioural research in arithmetic (Ferritor et al., 1972; Marholin & Steinman, 1977). For example, Marholin and

Steinman (1977) found a decrease in response rate when contingencies were placed on accuracy alone. Glynn's (1981, 1982a) commentary is particularly relevant to the procedures adopted in Experiment 1, since the teacher complained of a very low writing rate by her slow learners. Everything she had done to modify the problem had failed. Before it was possible to modify various writing aspects (e.g., creativity, word usage), it was necessary to obtain writing per se.

Interventions have tended to consist of consequential procedures such as token reinforcement, oral and written teacher comments, and public posting of performance. As a result of this emphasis, the value of antecedent stimuli or setting events (Glynn, 1982b) has been ignored (Kazdin, 1981). Singh et al. (1984) suggested that there was little research on the influence of antecedent stimuli on academic behaviour. Experiment 1 used correspondence training, a procedure that focussed on both antecedent and consequent stimuli.

The correspondence training paradigm

Correspondence training is a procedure for developing a relationship between an individual's verbal cues and their corresponding target behaviours. Israel (1978) differentiated between positive and negative correspondence. The majority of research focussed on positive correspondence, commonly referred to as simply correspondence. Positive correspondence was defined as the presence of both the verbal and nonverbal behaviours (saying X and doing X), when a person promised to perform a future behaviour (Paniagua & Baer, 1982). Negative correspondence was defined as the absence of both the verbal and nonverbal behaviour (not saying X and not doing X), when

a person reported about a prior behaviour (Paniagua & Baer, 1982). Karlan and Rusch (1982) made a further distinction, that of noncorrespondence. This was defined as when a person did not do what he said he would, or when he reported having done something that in fact was not done (saying X and not doing X) or when an act was performed but neither promised nor reported (not saying X and doing X).

Applying these concepts to the three phases of correspondence training, Karlan and Rusch (1982) suggested that during baseline one would expect a high proportion of negative correspondence with some noncorrespondence, and little positive correspondence. If noncorrespondence occurred during verbal training (an optional phase) the correspondence training phase was introduced. If this was effective, positive correspondence should increase and noncorrespondence decrease.

Risley (1977) has suggested that the essential components of the procedure include description of the behaviours required in the situation; comment by the audience on the correspondence between verbal and nonverbal behaviour; predictable occurrence of training sessions; guidance and reinforcement of the verbal statements such that they modified the desired target behaviour, rather than behaviours already in the repertoire; and feedback to the audience so that they knew the true performance in the setting where the behaviour was performed, thus providing accurate commentary on the correspondence relationship.

Investigations of a generalized relationship between verbal and nonverbal behaviour began in the early 1960s, with the focus on preschool children. The initial studies

were influenced by Luria (1959, 1961) and Vygotsky (1962). They observed that children often talk to themselves, and proposed that this speech-for-self functioned to orient or direct their behaviour. Luria (1959, 1961) suggested that as the child became articulate, his/her ability to respond to external verbal commands, to ascribe a verbal command's content to a signal, and to internalize verbal commands, increased. As a result, early studies set out to demonstrate that self-instructions facilitated corresponding nonverbal behaviour, specifically lever pushing (Bem, 1967; Birch, 1966) and finger tapping (Meichenbaum & Goodman, 1969a, 1969b). Others demonstrated that appropriate responding was facilitated and inappropriate responding impeded by training a child to self-instruct (Meichenbaum & Goodman, 1971; Monahan & O'Leary, 1971; O'Leary, 1968). It was only with the insight that a learning procedure may be necessary to achieve verbal self-control that correspondence training was developed.

With regard to verbal training or conditioning, reinforcement procedures were applied solely to the modification of verbal behaviour. Lovaas (1961, 1964) focussed on aggressive acts, rate of verbal behaviour and food preferences; Sherman (1964) on attention to toys; and Brodsky (1967) on the social behaviour of mentally retarded adults. The results were mixed, with little increase, if any, in corresponding nonverbal behaviour. Israel and Brown (1977) investigated the role of verbal training with preschoolers and play materials (plastic dinosaurs and cars) and concluded that it was a control procedure and an unnecessary precursor to correspondence training. Hence verbal control of nonverbal

behaviour could be achieved by correspondence training alone, without the preceding verbal training phase to strengthen the verbal component.

Early research investigated factors influencing correspondence. However, Risley and Hart (1968) were the first to develop a training procedure which successfully established generalized correspondence between verbal and nonverbal behaviour. They focussed on the blocks and painting behaviour of preschoolers and demonstrated that after repeated reinforcement of the correspondence, reinforcement of verbal behaviour alone could increase nonverbal behaviour, and hence maintain the correspondence.

Unlike studies in the 1960s which presumed a corresponding relationship between verbal and nonverbal behaviour, the research of the 1970s focussed on developing functional correspondence which was not apparent, and examined the processes which led to correspondence being achieved. Israel and O'Leary (1973) subsequently found that reinforcing a verbal-nonverbal sequence was effective. Reinforcing verbal behaviour alone led to only small increases in the corresponding nonverbal behaviour. They also worked with preschoolers, this time with activities (puzzles, letters, beads) in a play period. However, comparison between the two studies was impeded by their opposite correspondence sequence, and by Israel and O'Leary's (1973) failure to report direct measures of positive correspondence. Also subjects in the latter study trained for eight days with a strange experimenter, in contrast to Risley and Hart's (1968) subjects who trained for 28 days with the classroom teacher. Researchers currently adopt the say-do sequence. Israel and

O'Leary (1973) suggested that this sequence produced quicker correspondence because encouraging the child to say before doing might facilitate the use of this still novel regulatory function of language. As well, verbal behaviour was a more readily available and versatile discriminative stimulus than nonverbal behaviour. As such, it was more likely to prompt rehearsal and thus regulate or direct behaviour.

Others have also compared the do-say and say-do sequences. Israel (1973) focussed on play activities with preschoolers and found do-say correspondence slightly superior to say-do. However, they suggested this difference was due to methodological limitations. Conversely Karoly and Dirks' (1977) results on performance in a self-control tasks with preschoolers supported Israel and O'Leary's (1973) findings. They explained the higher level of correspondence with say-do over do-say groups in terms of the acquisition of a higher-order response set facilitated by the logic of an intention-execution sequence. Kanfer and Karoly (1972) offered an associative chaining view, in which the say-do group, unlike the do-say, had a verbal discriminative stimulus to cue subsequent performance. Alternatively Rogers-Warren and Baer (1976) hypothesized that the effects were due to delayed reinforcement of nonverbal behaviour, rather than the usual interpretation of reinforcement of true verbal statements. Redd (1969) interpreted the effects as due to observers acting as discriminative stimuli for the child's appropriate or good behaviour.

Rogers-Warren and Baer (1976) considered that the say-do and do-say sequences were based on the same procedural approach, as both influenced the next opportunity to perform

the target behaviour. They differed only in the timing of the opportunity to perform, immediately for the say-do and approximately 23 hours later for the do-say. Thus do-say was predicted to have a greater range of control, and hence was more likely to lead to generalization. Israel (1978), however, cautioned against such a conclusion citing existing support for the say-do procedure (Israel & O'Leary, 1973; Karoly & Dirks, 1977).

A different interpretation of this issue has been provided by Paniagua and Baer (1982) who replicated the major findings of other studies using similar procedures. However they questioned whether Luria's (1961) regulatory function was the reason for the superiority of the say-do over the do-say sequence. Alternatively they suggested that more fundamental than sequence per se, was where in those training sequences the reinforcer was placed. Sequence may be important only in determining the behaviour which will precede the reinforcer most closely. They conceived of correspondence training as a chain consisting of (a) a promise; (b) a series of behaviours leading from the promise to its nonverbal fulfilment, called intermediate behaviours; (c) a nonverbal behaviour that fulfils the promise; (d) a verbal report; and (e) a nonverbal (reported) behaviour. In their three experiments with normal preschoolers, participation was always highest when reinforcement was contingent on either the promise or the intermediate behaviours and delivered contingent on fulfilment of the promise; rather than when only the promise or only the report of nonverbal behaviour was directly reinforced. They explained this by saying that in the latter procedure, reinforcement was programmed to

occur only at one point. The total chain was better controlled in the former contingency, as reinforcers were operating at two points in the chain. Thus they concluded that the say-do (or promise-fulfil) sequence was more appropriate to train in the latter condition where the reinforcer was contingent on both the promise or immediate behaviours and on the fulfilment of the verbal behaviour or promise. However they suggested it was inappropriate to give a reinforcer for a promise regardless of the fulfilment of the promise.

The application of correspondence training

Despite its recent innovation, the correspondence training paradigm has been applied to a wide range of behaviours.

Within an adult population, Tracy, Bridell and Wilson (1974) assessed the effect of behavioural control of two classes of behaviour in a multiple-baseline design. Statements about activities generalized to participation in activities, but generalization of verbal responses about people did not occur in another stimulus setting. Unlike studies with children, which stressed that correspondence must be reinforced, the results of this study with a chronic psychiatric population were achieved by reinforcement of the verbal behaviour alone. The authors attributed this to group pressure which may have assisted in promoting correspondence, and the fact that in an adult population some verbal-nonverbal correspondence could be assumed. Whitman, Schibak, Butler, Richter and Johnson (1982) also made this point when they suggested that it may be possible simply to reinforce verbal behaviour in order to increase corresponding nonverbal behaviour, for children whose reinforcement

histories included reinforcement for matching verbal and nonverbal behaviour. Empirical support for this comes from Burron and Bucher's (1978) study with children.

Crouch, Rusch and Karlan (1984) successfully utilized the correspondence training paradigm to improve the work performance of three mentally retarded adults. The amount of time these adults spent on target tasks decreased and the accuracy with which they initiated tasks on time increased, as a result of their job-related time statements being reinforced. As with Tracy et al. (1974) the verbal training phase of the correspondence training paradigm alone was sufficient to establish and maintain verbal control of the required behaviour.

With regard to school-to-home generalization, Jewett and Clark (1979) reported that family meal conversation skills at home could be modified through implementation of correspondence training in a preschool situation. Baer, Osnes and Stokes (1983) also chose to train correspondence between verbal behaviour at school and nonverbal behaviour at home, since most studies focussed on verbal-nonverbal behaviours within the same setting. Like Glynn (1981) they perceived an advantage of correspondence training to be the possibility of controlling behaviour in remote situations, through control of verbal behaviour in more accessible, convenient settings. This enhanced the procedure's educational and clinical utility. Baer et al. (1983) chose more socially useful behaviours than toy playing, namely picking up pyjamas and clothes and choosing fruit for dessert. Delayed reinforcement (Fowler & Baer, 1981) was applied to correspondence training, as it has successfully led to generalization (Jewett & Clark, 1979;

Whitman et al., 1982). Baer et al. (1983) found correspondence across the two settings and generalization through delayed reinforcement of verbalizations.

Within the school setting, Whitman et al. (1982) applied correspondence training procedures to several unique areas. They were the first to focus on low IQ and mentally retarded children. Research had demonstrated that mentally retarded children could use verbal strategies if trained (Borkowski, Cavanaugh & Reichardts, 1978). Unlike the usual focus on preacademic and prosocial behaviours, they chose common classroom problems (out of seat behaviour, sitting posture, on task behaviour). Thirdly, they considered maintenance over time, generalization over situations and transfer to classroom teacher, issues which the correspondence training research had not consistently addressed. Results from their three experiments provided evidence for correspondence training's effective use with mentally retarded children, as well as support for their modification of the say-do to a show-do procedure with nonverbal children.

Rogers-Warren and Baer (1976) applied correspondence training to the prosocial behaviours of preschoolers' sharing and praising. Thus they extended previous research from simple to complex behaviours and from nonsocial to social behaviours. Correspondence developed and they proposed the procedure as an efficient means of increasing social behaviours. In a further study, Rogers-Warren, Warren and Baer (1977) systematically analyzed five components of the procedure used in the former study (Rogers-Warren & Baer, 1976). The results replicated those of the earlier study, finding the combined package of modelling, self-reporting and

reinforcement for true reports of sharing, to be the most effective.

Ballard and Jenner (1981) applied the correspondence training procedure to the social interactions and play activities of two primary school children with low social interaction levels. The two target behaviours increased and results were maintained at nine weeks' follow-up. Finally, Risley (1977) has demonstrated the effectiveness of the say-do paradigm which he applied to preschoolers' picking up of rubbish and polite verbal statements to the teacher.

Despite its wide application, the correspondence training paradigm has rarely been applied to children's story writing behaviours. Christie and Ballard (1983) applied the procedure to writing more sentences with normal seven-year-olds. The procedure was preceded by a setting event of having four charts to demonstrate the importance of number of sentences, different describing words, different action words and other techniques for effective storywriting. The say-do correspondence procedure phase showed a small and inconsistent increase in storywriting responses. Story quality did not improve between baseline and intervention phases. However the teacher could effectively and consistently carry out the procedure.

Mander and Monsen (1985) applied a writing package consisting of academic training, homework tasks and say-do procedures to the writing behaviours (number of sentences, describing and action words) of three problem primary school boys. All three target behaviours increased. There was a minor improvement in story quality and children's attitudes towards self and writing. However the intervention lasted

only two weeks with no maintenance or follow-up assessment.

Harris and Graham (1985) applied a self-control, rather than teacher-controlled, training approach to the composition skills (number of different action words, action helpers, and describing words) of learning disabled students. The treatment package consisted of strategy training, self-regulation training and instruction in the significance of these activities. Two of the four self regulation training components (self-determined criterion-setting and self-assessment) were similar to the correspondence training procedure. The various parts of speech, and mean number of words per story increased in intervention. These stories were also rated as being of higher quality than baseline stories. Maintenance and generalization probes, until 14 weeks post training, were positive.

Advantages of the intervention

Correspondence training has been demonstrated to be an effective technique for improving social and work skills, classroom problems and story writing behaviours. The procedure has other advantages. Two suggested by Israel (1978) were the accessibility of a subject's verbal behaviour, and the maintenance of nonverbal target behaviour without direct monitoring and reinforcement. For these reasons, Whitman et al. (1982) considered the procedure more efficient and effective in modifying class behaviour than reinforcement procedures which focussed on modifying nonverbal behaviour. They also pointed out its potential use with both adaptive and maladaptive behaviours and academic problems, in pre- and

elementary school as well as in the area of special education. Christie and Ballard (1983) suggested it provided a naturalistic intervention and was efficient in terms of teacher time and training. Whitman et al. (1982) supported this view, stating that its potential use by paraprofessionals in various settings was good because of its easy application and simplicity. They suggested it may be more valuable than the usual reinforcement interventions which were time consuming to implement and maintain. Despite the advantages, however, the authors pointed out that more research in various areas was required as well as analyses into its process and key components.

Feedback is a component of correspondence training. It is a consequating discriminative stimulus providing information about how a person performed on a task. In Experiment 1 the correspondence between verbal and nonverbal behaviours was reinforced in two ways. Firstly social reinforcement in the form of verbal praise, rather than tangible reinforcement, was used. This was chosen since schools operate verbally and do not have time to apply tangible reinforcers, which tend to be less appropriate with older pupils.

Performance-feedback was the second way the correspondence was reinforced. It was implemented by asking the writer to count the number of words written. Performance feedback is supported by laboratory (Barringer & Gholson, 1979; Kulhavy, 1977) and applied research (Jerram, 1985; Scriven & Glynn, 1983; Van Houten, 1979). Scriven and Glynn (1983) have argued that feedback can be excessively delayed and extremely infrequent in secondary schools. Regular,

frequent and consistent feedback was advantageous to a student's ability to monitor changes in performance across time. This would seem especially important with underachieving students. They had a history of failure experiences, which Bandura's (1977) model of self-efficacy predicted would lower self-efficacy and possibly self-esteem. In receiving immediate feedback as to the nature of the correspondence, hopefully positive, via the correspondence training paradigm, the underachieving student would experience mastery over a level of behaviour, that he/she both determined and fulfilled. According to Bandura (1977) performance accomplishment was the most successful way of influencing and raising a person's sense of efficacy. The concept of self-efficacy also applies to a point made by Macmillan and Forness (1970) and Kazdin (1981) who suggested that the target behaviour must come under the control of natural reinforcers in maintenance as soon as possible. During the maintenance process the desired behaviour was increased by natural consequences intrinsic to completion of tasks, feelings of self-worth, social approval, and the satisfaction of assuming self-responsibility. The process was conducive to developing self-control and independent learning skills, lacking in learning disabled students.

Maintenance and generalization

A final advantage of correspondence training pertains to the areas of maintenance and generalization. Maintenance of correspondence refers to the subject continuing to fulfil the promise, even though the correspondence is no longer reinforced. Generalized correspondence refers to the

subject's verbal promise controlling behaviour for which correspondence was not trained. Behavioural researchers have been criticized for their lack of attention to these areas (Agras & Berkowitz, 1980; Hayes, Rincover & Solnick, 1980). However in recent years, techniques have been developed to achieve maintenance and generalization (Kazdin, 1982; Marholin, Siegel & Philips, 1976; Stokes & Baer, 1977). Baer, Wolf and Risley (1968); O'Leary and Drabman (1971); and Russell (1974) found that programming was necessary to achieve generalization. Wildman and Wildman (1975) have suggested 15 rules for the promotion of generalization.

Ballard (1983) suggested that correspondence training was a useful self-management procedure for securing generalization and maintenance. Stokes and Baer (1977) even suggested its potential as a mediated generalization procedure to apply to the latter steps of a programme, once the new behaviour had been established, to ensure its performance in the appropriate, nontraining setting. Hops, Walker and Greenwood (1979) successfully carried out this procedure. Karlan and Rusch (1982) also suggested that correspondence training could be applied to existing naturally developed or trained target behaviours, to achieve maintenance and generalization.

Transfer of training was successfully achieved in a study by Whitman et al. (1982). The maintenance procedure was described as easy to administer by the teacher.

Baer, Williams, Osnes and Stokes (1984) programmed generalization and maintenance, using a delayed reinforcement technique, successfully used by others (Baer et al., 1983; Fowler & Baer, 1981; Whitman et al., 1982). They explained

the strategy by suggesting that postponing reinforcement until both verbal and nonverbal behaviours occurred made it very difficult to discriminate whether reinforcement was contingent only on verbal behaviour, or on the verbal-nonverbal correspondence. They also recommended that generalization which was not maintained could be recovered by briefly reinstating the original training contingencies. These and most studies (Israel & Brown, 1977; Williams & Stokes, 1982) have all achieved generalization, maintenance or both. There have been only a few exceptions (Israel & O'Leary, 1973; Karoly & Dirks, 1977). Three studies (Risley & Hart, 1968; Rogers-Warren & Baer, 1976; Williams & Stokes, 1982) returned to baseline after correspondence training, so as to assess maintenance. Not surprisingly, positive correspondence decreased, as maintenance was not programmed.

Although the research had shown some consistencies, further elucidation was needed (Baer et al., 1984; Whitman et al., 1982). Karlan and Rusch (1982) reinforced this view, when they called for more research into the environmental variable(s) that promoted maintenance, as well as how to achieve maintenance through the withdrawal of the verbal behaviour or saying. The latter question remains unexplored. They also recommended more research on generalized verbal correspondence since this seemed significant to the development of self-control. The issue of programming generalized verbal control across dissimilar behaviours has not yet been resolved. At present, generalized verbal control can only be effected across similar behaviours (Israel & Brown, 1977; Risley & Hart, 1968; Rogers-Warren & Baer, 1976; Williams & Stokes, 1982).

Despite the gains made thus far in the areas of maintenance and generalization, it must be concluded that more research is needed into correspondence training.

The application of correspondence training to story writing was the focus of Experiment 1. This study was unique in applying correspondence training to a group of high school boys who were categorized as slow learners. It was also unique in using a changing-criterion design, and in monitoring 23 collateral behaviours.

Experiment 1 examined five experimental questions:

(1) Did correspondence training increase the target behaviour - production of a previously set mean number of words?

(2) Were there improvements in the collateral behaviours?

(3) Did baseline essays improve qualitatively in intervention, as assessed by holistic scoring?

(4) Could control be transferred from experimenter to teacher without loss of intervention gains?

(5) Were target levels sustained in the maintenance phase?

METHOD

Subjects and Setting

The study took place in a third form alternative education classroom of a central city boys' private, Catholic secondary school. Being in this special class involved separation academically from the mainstream boys. The small class size afforded greater individual attention from the teacher. A mixture of slow learner programmes and alternative

subjects were taught. The aims of the class were to take pressure off the boys academically and socially and develop their self-esteem and confidence. Emphasis was placed on developing social, and basic academic and general life skills. This would lead in form four and five to Work Experience. Some vocational skills were also taught (e.g., workshop skills). Considerable assistance was provided by support services, e.g., psychiatric, health, and remedial teaching.

Placement in the alternative education class was determined by TOSCA and PAT tests and social maladjustment which usually occurred because of low academic ability. Academically they were two stanine or below in TOSCA and in the bottom 10% in PAT. Further details of subjects can be found in Table 1. No subjects were on medication.

Insert Table 1 here

The study was conducted daily, during the English period. Data were obtained from four white male students, although all children in the class received treatment procedures. The female teacher volunteered to participate in the study during transfer and maintenance phases. The only other person involved in the study was the experimenter, who implemented baseline and intervention phases, collected data, and trained and supervised the teacher during the transfer and maintenance phases.

Stimulus Materials

The boys were required to write in response to a stimulus prompt in the form of a story poster. These posters, 50 x 100 cm, were obtained from various organizations and supplemented the teacher's own supply. The poster, chosen by a different boy each session, was always new. The topics of these posters are presented in Appendix A.

Table 1

Subject details in Experiment 1

		Neale analysis of reading ability		PAT reading comprehension			PAT reading vocabulary			PAT mathematics		
Subjects	Age (yrs,mths)	Reading accuracy (yrs,mths)	Reading comprehension (yrs,mths)	Raw Score	Level	Percentile	Raw Score	Level	Percentile	Raw Score	Class Percentile	Age Percentile
Mark	13,6	8,11	9,3	11	5b	6	23	5b	10	9	2	2
Kevin	13,5	8,10	9,1	17	6a	17	24	5a	15	16	21	25
Martin	14,4	9,8	10,11	10	5c	4	30	6b	14	11	5	2
Brendon	13,8	9,1	9,3	2	2a	1	24	5a	11	14	14	16

Performance Variables

The correspondence training paradigm was applied to the target behaviour of producing number of words. This target behaviour was chosen since the classroom teacher was having difficulty getting her pupils to write and she requested programming assistance. Correspondence was defined on a session-by-session basis. It occurred when the subject set a realistic target number of words to be written, and then reached that target to meet the criterion for that day.

To quickly obtain a measure of the amount written per session, so consequences could be delivered before the session ended, the subject counted the number of words written and wrote the figure at the end of the essay. The accuracy of this figure was later checked by the experimenter.

In addition, measures were taken of collateral behaviours to assess whether these were indirectly affected by the intervention. These measures are discussed in the section on dependent variables.

Experimental Design

A changing-criterion design (Hartmann & Hall, 1976) was used to evaluate the effects of correspondence training on the written composition of four subjects. Following baseline, correspondence training was introduced. The level set for the first treatment phase was approximately 15% above each subject's mean number of words written during previous baseline sessions. The initial level, and proceeding levels, were negotiated individually with each subject.

Because it was important to set a level the subject could initially reach, it was lowered if the initial level was found to be too high. Intervention was followed by a transfer phase involving a shift in implementation from the experimenter to the teacher. The study concluded with a maintenance phase involving reinforcement only for initial verbal statements of the number of words to be written.

Whether the level was reached was determined in every session. If performance consistently reached the level for three consecutive days, the level was increased in consultation with the subject, so he decided the level of the next subphase of treatment. As performance stabilised at this new level, the level was again shifted upward. The level was altered until the desired ceiling was reached. When the target behaviour increased in a stepwise fashion in correspondence with the levels set, experimental control was demonstrated.

Procedure

Baseline Sessions were held daily during the usual 40-minute English period. At the beginning of each baseline session, a class member chose a story poster and the experimenter announced to the entire class: "You have half an hour to write an essay about the poster." No discussion or interaction between the experimenter and subjects took place. Each subject wrote down the number of words written at the end of the half hour, and this was verified by the experimenter who provided feedback to the subjects.

Correspondence training During this phase, each subject was asked at the beginning of the session to nominate the number of words he would write that day. When this

required number of words had been written for three consecutive days, the experimenter said to the subject: "You wrote x number of words. That is terrific but there is still room for improvement." The subject was then asked to nominate a new, higher number of words.

At the end of the half-hour session, when the number of words written had been counted, one of two alternative statements was made. If the subject met the number set for that day, the experimenter said: "You said you were going to write x number of words and you really did. I hope you can do as well tomorrow", and the subject was socially reinforced with smiles and praise. If the level for that day had not been met, the subject received the following feedback: "You said you were going to write x number of words, but you only wrote y words, didn't you? Well I hope you will do better tomorrow by reaching your target number of words."

Transfer In this phase, the transfer of control was shifted from the experimenter to the regular classroom teacher. The procedure the teacher conducted was identical to that in the intervention (correspondence training) phase. The transfer phase was terminated when all subjects reached one required level each with the teacher in control. The experimenter's role in this phase was to instruct the teacher on how to conduct the procedures and observe that she was doing this consistently.

Maintenance The procedure for this phase was similar to the transfer phase, except that reinforcement was no longer contingent on correspondence between the number of words nominated and the number of words actually written. The only reinforcement given was for the verbal statement

about the number of words the subject said he would write. No contact was made with the subjects at the end of the session, irrespective of whether the target was reached or not. At this stage the experimenter only visited the school to observe that the teacher was consistently implementing the procedure and to pick up the completed essays for analysis.

Dependent Variables

In Experiment 1 measures of both the target behaviour (number of words written) and collateral behaviours were taken. Each essay was analyzed according to these measures, and summarized on the record sheet in Table 2.

Insert Table 2 here

Mechanical component This component was assessed by the handwriting score (cursive). Based on the Test of Written Language (TOWL) (Hammill & Larsen, 1983) the following rules and instructions were used.

(a) Only score those samples of written composition, on this measure, which have been written in cursive. Thus printed essays are excluded. Sometimes printing and writing will be mixed up. In those cases, a minimum of 50% of total words must be written for the sample to be rated as cursive.

(b) Match as closely as possible the cursive handwriting in the story, with one of the five samples in Figure 1.

Insert Figure 1 here

Table 2

Record Sheet for Experiments 1 and 2

For each sample fill in the appropriate information and obtain the following measures.

NAME:
SCHOOL:
DATE WRITTEN:
TYPE OF STIMULUS PROMPT:
PHASE OF EXPERIMENT:
SCORER'S NAME:

COMPONENT		MEASURE	SCORE
Mechanical		Handwriting score (cursive)	- - -
Productive		Total no. of words written	- - -
		Total no. of sentences	- - -
		Total no. of thought-units	- - -
		Percentage of new words used	- - -%
		Vocabulary diversity score	- - -
Conventional		Percentage correct for capitalisation	- - -%
		Percentage correct for punctuation	- - -%
		Percentage spelling inaccuracy measure	- - -%
Linguistics		Vocabulary score	- - -
	Semantics	Word usage score	- - -
		Percentage mature word usage	- - -%
		Grammar errors per 100 words	- - -
		% total nouns	- - -%
		% total A. verbs	- - -%
		% total adjectives	- - -%
		% total adverbs	- - -%
	Syntax	% diff. nouns	- - -%
		% diff. A. verbs	- - -%
		% diff. adjectives	- - -%
		% diff. adverbs	- - -%
		mean T-unit length	- - -
Creative		Thematic maturity score	- - -
		Creativity score	- - -
Contact		Percentage of words asked how to spell	- - -%

FIGURE CAPTION

FIGURE 1. Scoring guide for cursive handwriting. From The Test of Written Language (p.23) by D.D. Hammill and S.C. Larsen, 1983, Austin, T.X.: PRO-ED.

Rating 10

Once upon a time there was a Mars club around the corner where every day most

Rating 9 of the people in my class went to after school at three o'clock every. Soon later on in the year five people had joined. Soon

Rating 8

The first picture is about sciencetist have notice that the earth is start-

Rating 7 ing to explode. Because the earth is cracking. They got all of the people together because they have notice that there

Rating 6

Once upon a time in a galaxy far away on the planet of Qunbee. Qunbeeen Shaken

Rating 5 by a series of Qunbee quacks. This planet has a very well advanced science department it claims the planet will be

Rating 4

Once upon a time some space ships landed on some planet were there was

Rating 3 elephants, mice, ducks, turtles, pegs and hipptalpatimos and the people started digging holes in the the

Rating 2

onet a time that was is popued fonae meme and the tirz and tiry

Rating 1 to git the aila form oat paped and the was not art con foune poposed they was not

Rating 0

(c) The main consideration is legibility. Factors such as slant, spacing, letter formation and size should not be emphasized, and are important only in their contribution to or adverse effect on readability.

(d) No allowance is to be made for the age or experience of the student. Use the absolute criterion, the samples, when judging the handwriting.

(e) Consider only cursive handwriting and ignore such aspects as spelling errors, word usage, grammar errors and literary merit.

(f) Rate the handwriting on a scale of 0 to 10. Avoid assuming that a 5 is average.

(g) Before scoring experimental samples of written composition, obtain practice using the handwriting samples in Hammill and Larsen (1983, pp.90-92) so you are proficient. Throughout evaluation review this practice task if you feel unconfident or out of practice.

Productive component Five measures were assessed within this component.

(a) Total number of words written. A word was defined as any group of letters, written or printed, representing a spoken word (Maloney & Hopkins, 1973). This measure provided an estimate of written output. Scorers were instructed to count the total number of words written, according to the following rules:

1. Disregard titles, e.g., The End.
2. Hyphenated words, e.g., bow-wow are treated as single words.
3. Contractions, e.g., didn't, are considered as one word.

4. Exclude numerals, e.g., 7.00; 8; and abbreviations, e.g., a.m., USA.
5. Words do not have to be spelt correctly to be included.
6. Omit a series of letters joined together, but unrecognizable as a word, that is, meaningless language or incomprehensible words.

(b) Total number of sentences. A sentence was defined as "beginning with a capital letter and/or on a new line, and/or having a period, question mark, or exclamation point at the end, and/or containing at least one subject and predicate." For cases where capital letters and periods occurred infrequently, a sentence was defined as "a group of words that made sense as a sentence" (Ballard & Glynn, 1975, p.389). Only complete sentences were counted.

(c) Total number of thought-units (T-units). A T-unit was defined as "a segment of meaningful expression that contains an identifiable verb and its subject and that can stand alone, that is a complete sentence" (Hammill & Larsen, 1983, p.44). It was not a fragment but conveyed a complete thought. Scoring instructions included:

1. Mark the end of each T-unit with a slash (/). Count up the number of slashes in the story, to determine the total number of T-units.
2. Ignore incorrect punctuation, capitalization and spelling, and mark each unit as if it were correctly written.

3. Do not count a T-unit that is incomprehensible, even if it has a verb and subject and stands alone.
4. To get practice in measuring T-units, refer to Hammill and Larsen (1983, p.45). Become proficient before you start analyzing experimental samples of written composition.

(d) Percentage of new words used. A new word was one the student had not used before in any previous stories. Scorers were instructed to make an alphabetical list of all the words used, starting with story one, to refer to when assessing the number of new words used in each story. The procedure used was:

1. Beside each word write the date it was written and whether the verb or noun form of the word was used, if it can act as two different parts of speech.
2. Correct the spelling if necessary but do not change the form of the word so much that in doing so a new word is made which the subject may not be capable of creating.
3. Do not count the plural version of a word already counted as new in the singular form, and vice versa.
4. Count up the total number of new words and divide it by the total number of words written and multiply by 100 to get the percentage of new words used.

(e) Vocabulary diversity score. This measure was concerned with the number of words within each story that

were different. A different word was defined as the first usage in that day's story (for example 'I like strawberries, I like pineapple', totals six words but only four different words). This gave a measure of whether the writing consisted of repeated words or different words. Instructions for this measure included:

1. Do not count a plural version of a singular word as different. (For example chair/chairs = one different word).
2. If the same word appears within the same essay, but used in different forms, then count as different words. For example Give me a go (go = noun) and Please go (go = verb). Count go as two different words.

The vocabulary diversity score was calculated using a formula designed to control for the decrease in proportion of different words with increasing total words (Carroll, 1964). The formula measured the number of different words (types) used, independent of the total number of words (tokens) in the written sample. The formula was $Nd\sqrt{2Nt}$, where Nd = number of types and Nt = number of tokens. The advantage of this formula over a simpler percentage measure is demonstrated in the following comparison, where all four total words are different (example one) and where all 30 total words are different (example two). With the percentage method, whether four or 30 different words are used, the total is still 100%. However, the formula clarifies the situation, by demonstrating that writing 30

different words within a total output of 30 words (vocabulary diversity = 3.8) is a lot more difficult than writing all different words when a total of only four words are written (vocabulary diversity = 1.4).

Conventional component This component was concerned with assessing the boy's ability to write in compliance with rules governing capitalization, punctuation and spelling. Three measures were assessed.

(a) Percentage correct for capitalization. To calculate this measure the scorer tallied up the number of letters correctly capitalized and divided this by the total number of letters which should have been capitalized plus the number of letters which were inappropriately capitalized (Bording et al., 1984). This figure was then multiplied by 100 to obtain the percentage correct for capitalization.

The scorer was provided with a chart of capitalization errors to guide scoring this measure. The chart was a modified version of that devised by Burns (1974) as seen in Table 3.

Insert Table 3 here

(b) Percentage correct for puncutation. This was calculated by tallying up the total number of correct punctuation marks and dividing this by the total number of appropriate punctuation marks necessary plus the number of incorrectly used punctuation marks (Bording et al., 1984). This figure was then multiplied by 100 to obtain the percentage correct for punctuation.

Table 3

Chart of Capitalization Errors

Pupil's name:

Date written:

Letters requiring capitalization										
First Word of sentence	The word "I"	Proper names	Title	Proper adjectives	First word of quotation	Words for the deity	Appropriate abbreviations	Salutation & closing of letter	Words such as mother, father when used as a name	Trade names

Total errors =

Note. From Diagnostic teaching of the language arts (p.90) by P.C. Burns, 1974, Itasca, Illinois, F.E. Peacock Publishers.

The scorer was assisted by using a modified chart of punctuation errors, originally developed by Burns (1974) as seen in Table 4.

Insert Table 4 here

(c) Percentage spelling inaccuracy measure. This was derived by counting up all spelling errors, including the same word misspelt more than once, and dividing this by the total number of words in the story. This figure was then multiplied by 100 to obtain the percentage of words spelt incorrectly. The raw data were converted to percentage measures to control for increasing total word output. This allowed true changes in proportion to be detected rather than changes due to increased output.

Linguistic component This component consisted of both semantic and syntactic measures.

(a) Semantic measures

1. Vocabulary score. This measure involved counting the number of words in the composition that consisted of seven or more letters. The following guidelines provided by Hammill and Larsen (1983) were used for scoring:

COUNT

i. Made-up words of seven or more letters, e.g., moontroplis.

ii. Misspelled words which if spelled correctly would consist of seven or more letters, for example Janary (January).

Table 4

Chart of Punctuation Errors

Pupil's Name

Date written:

Punctuation Errors								
Period	Comma	Semi-colon	Colon	Quotation mark	Apostro-phe	Question mark	Hyphen	Under-lining

Total errors =

Note. From Diagnostic teaching of the language arts (p.90) by P.C. Burns, 1974, Itasca, Illinois, F.E. Peacock Publishers.

iii. Words correctly hyphenated, for example ill-natured.

iv. One long word written as two short words, e.g., foot ball (football).

v. Repetitions of words of seven or more letters.

DON'T COUNT

i. Words inappropriately hyphenated, for example shirt-collar.

ii. Abbreviations or acronyms, even if seven or more letters long.

iii. Spelling errors that are seven or more letters in length when not so if spelled correctly, e.g., travell (travel).

iv. Numerals, e.g., 1,000,000

2. Word usage score. This was derived from a simple rating scale to assist the scorer in subjectively assessing the subject's written vocabulary. The subject's written composition was given a ranking of either 3, 2, 1, or 0, based on the guidelines in Table 5, to derive a word usage score. The

Insert Table 5 here

guidelines were a modified version of those provided by Burns (1980).

3. Percentage mature word usage. Mature words were less common, infrequently used words and contrasted with undistinguished or immature, high frequency words. To obtain a measure of percentage mature word usage, the number of words

Table 5

Guidelines for the Word Usage Score

RANK	CRITERIA
3	The composition contains a variety of clear, precise, descriptive, vivid words that (1) appeal to the senses, (2) develop shades of meaning, (3) define action, (4) enhance word pictures, and (5) reflect effective similies and/or metaphors.
2	The words in the composition are adequately descriptive. There is sporadic use of vivid words or phrases.
1	There is little variety of word choice; very few descriptive or picture words; and common or overworked similies and metaphors.
0	Only trite, ineffective, dull and monotonous words are used.

Note. From Assessment and correction of language arts difficulties by P.C. Burns, 1980, Columbus, O.H., Merrill.

in the sample of written composition that did not appear in Table 6 of undistinguished word choices

Insert Table 6 here

(Finn, 1977) were counted. Repetitions and words spelt incorrectly were included. The total of mature words was divided by the total number of words written and multiplied by 100.

(b) Syntactic measures

1. Grammar errors per 100 words. Grammar errors were tallied according to the categories shown in Table 7. The total number of grammar errors

Insert Table 7 here

was then divided by the total number of words and multiplied by 100, which converted to grammar errors per 100 words.

2. Percentage total nouns. Categories of nouns included in this measure were common nouns (e.g., pen, pencil, pigsty); abstract nouns (e.g., wisdom, justice, equality); proper nouns (e.g., Asia, Churchill, Elizabeth) and nouns of material (e.g., wood, glass, wool) (Moffat, 1968). Personal pronouns (such as he/she) were excluded from the noun count. Words could often act as a noun, adjective, verb or adverb, depending on how they were used in the text. Scorers were instructed to verify word usage using a dictionary if they were unsure. The total number of nouns, including repetitions, was divided by the total number of words in the story and multiplied by 100.

Table 6

Undistinguished word choices

a	doesn't	house	no	take
able	doing	how	not	talking
about	done	I	now	that
again	don't	if	of	that's
air	down	I'm	off	than
all	drink	important	on	the
also	dumping	in	once	their
and	each	into	one hundred	them
animals	earth	is	open	then
another	either	it	or	there
any	else	its	other	they
anything	enough	job	our	thing
are	even	just	out	things
around	every	keep	over	think
as	everyone	kill	own	this
ask	everything	know	part	time
asked	fair	lake	pay	to
at	family	lakes	people	too
away	feel	let	person	try
bad	few	like	place	until
be	filter	little	plants	up
because	find	live	problem	us
been	first	living	put	use
before	fish	long	rather	very
being	fishing	look	reason	want
better	food	lot	right	was
big	for	make	river	water
build	form	making	run	way
business	from	man	running	we
but	get	many	said	well
buy	getting	may	same	what
by	give	maybe	save	when
call	go	me	say	where
called	going	mean	see	who
came	good	men	should	why
can	got	might	since	will
children	had	money	so	with
cleaning	happen	months	some	without
clear	happy	more	someone	won't
close	hard	most	something	work
closed	have	much	soon	worked
come	having	must	source	working
could	he	my	start	world
couldn't	health	near	stay	would
day	help	need	still	years
did	him	never	stop	you
didn't	his	new	support	your
do	home	next	swim	

Note. From P.J. Finn, Computer-aided description of mature word choices in writing. In Evaluating writing by C.R. Cooper and L. Odell, 1977, Champaign, Ill., National Council of Teachers of English.

Table 7

Grammar Analysis Chart

Pupil's name:

Date written:

VERBS				PRONOUNS			ADJECTIVES/ADVERBS			WORDS				
Tense form	Tense agreement	Agreement	Auxiliary missing	Subject or object position	Pronoun/ Adjective	Antecedents	Form	Confusion	Article comparison	Addition	Omission	Substitution	Plural	Substandard

Total errors =

Note. From Diagnostic teaching of the language arts (p.101) by P.C. Burns, 1974, Itasca, Illinois, F.E. Peacock Publishers.

3. Percentage total action verbs. Action verbs were defined as expressions of act, occurrence or movement but not a mode of being, as any form of the verb 'to be' (Ballard & Glynn, 1975). Auxiliary verbs helped the principal or main verb to form the exact tense and mode which the sense required, and needed to be distinguished from action verbs. The scorers were given the following instructions:

   i. Count only the action verbs, excluding all auxiliary verbs, as in the following example: For "He is playing", "We have driven", and "We will go", 'playing', 'driven' and 'go' are action verbs. 'Is', 'have', and 'will' are auxiliary forms of the verbs "to be" and "to have". Like other verb auxiliaries (will, would, shall, should, can, could, may, might) they should not be counted.

   ii. An exception to this rule is that auxiliary verbs may be included only when they act as the main verb. For example, "She is happy", "I have homework", and "Yes I will". Here 'is', 'have' and 'will' are counted as they are not being used in an auxiliary sense but as main verbs.

   iii. When two verbs occur together and the second is in the inifinitive (e.g., to fly) count both verbs, as in the example, "He is going to fly". 'Is' acts as an auxiliary to the verb "going" and so is not counted. 'Going' and 'fly' are counted.

Scorers were required to count up the total action verbs (including repetitions and different tenses of similar verbs) and divide this by the total verbs (including all action and main verbs) and multiply this by 100.

4. Percentage total adjectives. An adjective was defined as "a word serving as the modifier of a noun to denote the quality of the thing named, to indicate quantity or extent, or to specify a noun as distinct from something else" (Ballard & Glynn, 1975, p.389). Adjective types included descriptive (e.g., good, pretty); numeral (e.g., ten, first); possessive (e.g., my, our); interrogative (e.g., which book?); demonstrative (e.g., that book); distributive (e.g., each boy) indefinite (e.g., some boys; any boy); and proper (e.g., American cars) (Moffat, 1968). Others included articles (e.g., this, that, these, those; not a and the) and quantitative words (e.g., other, another, no). The total number of adjectives was counted including repetitions and divided by total number of words and multiplied by 100 to convert to a percentage of total adjectives.

5. Percentage total adverbs. An adverb was defined as "a word that modifies a verb, adjective, another adverb, preposition, phrase, clause or sentence and expresses some relation of manner of quality, time, place, degree, number, cause, opposition, affirmation or denial" (Ballard & Glynn, 1975, p.389). Examples of these forms (Moffat, 1968, p.37) were given to scorers.

Adverbs:

i. modify verbs (e.g., She fell in; The chimney fell over). The adverbs "in" and "over" modify the verb "fell". They should not be confused with prepositions in the following example (He fell in the pond; She came to Christchurch)

ii. modify adverbs (e.g., He talks too fast; He fell in badly)

iii. modify adjectives (e.g., delightfully slim; very beautiful).

Generally adverbs answered the question how? when? where? or why?, and were classified as adverbs of:

iv. manner (e.g., easily, badly, well)

v. time (e.g., now, yesterday, never)

vi. place (e.g., here, there, everywhere)

vii. reason (e.g., consequently, inasmuch)

viii. degree (e.g., very, too, tremendously

and interrogative adverbs in similar classes. For example, interrogative adverbs of:

ix. manner (how?)

x. time (when?)

xi. place (where?)

xii. reason (why?).

The total number of adverbs was counted including repetitions and divided by total number of words and multiplied by 100 to convert to a percentage of total adverbs.

6. The next four measures referred to percentage different nouns, action verbs, adjectives and

adverbs. These measures were obtained in an identical fashion to percentage total nouns, action verbs, adjectives and adverbs, except that repetitions were not included. Percentage different parts of speech were obtained in addition to percentage total parts of speech, so that a comparison of the two could indicate how much repetition was involved in the percentage total parts of speech measures. The different nouns measure did not include plurals. Thus, dog/dogs and men/man were treated as one "different" word not two. With regard to different action verbs, this category did not include a different tense of a verb previously used in the story being counted as a separate response. Thus "see" and "saw" were treated as one different action verb from the verb "to see".

7. Mean T-unit length. This index of syntactic complexity was calculated by dividing the total number of words written by the number of T-units.

Creative component This section consisted of two measures.

(a) Thematic maturity score. The ten items that formed this checklist (Table 8) were derived from the 20 items

Insert Table 8 here

that form the Thematic Maturity Score in the Test of Written Language (Hammill & Larsen, 1983, pp.21-4). The modification of these 20 items was necessary because in the Test of Written Language form they were developed for a specific set of three pictures (which

Table 8

Thematic Maturity Score

Pupil's name:

Date written:

Category	0 or 1
1. Writes in paragraphs	
2. Mentions any object shown in the picture	
3. Writes a story that has a definite ending	
4. Expresses some philosophic or moral theme	
5. Have a title	
6. Uses dialogue	
7. Attempts humour	
8. Writes an integrated story about the picture	
9. Relates themes/plots that aren't directly implied in the picture	
10. Attempts to develop the personalities/features of one or more characters/objects	

Total Score =

were to form a story). In the present study the picture topic changed in each session so the 20-item checklist was inappropriate. The modified checklist could be applied to all picture prompts in both experiments.

Instructions given to the scorers were:

1. Become familiar with the ten items of the Thematic Maturity Score. Their definition and scoring guidelines are provided in Hamill and Larsen (1983, pp.22-4).
2. After reading the composition, evaluate it by means of the ten items, giving either 0 or 1.
3. The subject's total score can range from 0 to 10. Thus a score of ten would mean that all items were included in a story.

(b) Creativity score. In order to assess the creative responses of each written composition, the Carlson Analytical Originality Scale (Carlson, 1961, 1963, 1965, 1973) was adopted (Table 9). Carlson (1961) defined

Insert Table 9 here

original writing as a form of writing which was individual, novel or unusual and expressed through the dimensions of story structure, novelty, emotion, individuality, and story style. Her 36-item scale covered these five dimensions and each item was rated on a 0 - 5 point scale. In the present study, the scoring was changed to a 0 - 3 scale for two reasons. Firstly because the sample populations did not write creative enough responses to warrant a 5-point scale. Secondly,

Table 9

Carlson Analytical Originality Scale Key (modified)

Pupil's name:

Date written:

Scale Division A - Story Structure		
1.	Unusual title	0 1 2 3
2.	Unusual beginning	0 1 2 3
3.	Unusual dialogue	0 1 2 3
4.	Unusual ending	0 1 2 3
5.	Unusual plot	0 1 2 3
Scale Division B - Novelty		
6.	Novelty of names	0 1 2 3
7.	Novelty of locale	0 1 2 3
8.	Unique punctuation and expressional device	0 1 2 3
9.	New words	0 1 2 3
10.	Novelty of ideas	0 1 2 3
11.	Novel devices	0 1 2 3
12.	Novel theme	0 1 2 3
13.	Quantitative thinking	0 1 2 3
14.	New objects created	0 1 2 3
15.	Ingenuity in solving situations	0 1 2 3
16.	Recombination of ideas in unusual relationships	0 1 2 3
17.	Picturesque speech	0 1 2 3
18.	Humour	0 1 2 3
19.	Novelty of form	0 1 2 3
20.	Inclusion of readers	0 1 2 3
21.	Unusual related thinking	0 1 2 3
Scale Division C - Emotion		
22.	Unusual ability to express emotional depth	0 1 2 3
23.	Unusual sincerity in expressing emotional personal problems	0 1 2 3
24.	Unusual ability to identify self with feelings of others	0 1 2 3
25.	Unusual horror theme	0 1 2 3
Scale Division D - Individuality		
26.	Unusual perceptive sensitivity (social and physical environment)	0 1 2 3
27.	Unique philosophical thinking	0 1 2 3
28.	Facility in beautiful writing	0 1 2 3
29.	Unusual personal experience	0 1 2 3
Scale Division E - Style of Stories		
30.	Exaggerated tall tale	0 1 2 3
31.	Fairy tale type	0 1 2 3
32.	Fantasy-turnabout of characters	0 1 2 3
33.	Highly fantastic central idea of theme	0 1 2 3
34.	Fantastic creatures, objects, or persons	0 1 2 3
35.	Personal experience	0 1 2 3
36.	Individual story style	0 1 2 3

Note. From Sparkling words: Two hundred and twenty-five practical and creative writing ideas by R.K. Carlson, 1973, Geneva, Illinois, Paladin House Publishers.

the reduction minimized subjectivity and thus inter-scorer differences, which tended to occur with a wide scale range.

The scorer was provided with a xeroxed section of a book (Carlson, 1973) which explained the scale and its development, gave directions for its use, and gave definitions and examples of points 0, 1, 3, 5 within each of the 36 items, which correspond to the modified version's points of 0, 1, 2, 3 respectively. Thus the finer distinctions of points 2 and 4 in the original scale were excluded in the modified version of the scale.

Holistic scoring

In order to achieve a comprehensive and valid assessment of intervention effects, analytic scoring was supplemented with holistic scoring.

A class of stage two education students doing a measurement course were given a packet of ten essays. Holistic scoring was obtained on only a portion of the total samples. The last eight essays of each boy's baseline and intervention phases were selected and photocopied four times. This was done so the same essay could be marked by four different people. The packets of essays were made up by randomly selecting a different baseline and intervention essay from each boy. The order of baseline and intervention essays was randomized.

The undergraduates were blind to the experimental protocol and did not know which essay related to baseline or intervention phases. The subject's name, date and other identifying factors were removed from the essay before xeroxing it,

to control for confounding effects of extraneous variables. Essays were not typed before scoring.

The scorers were given the following instructions:

(1) Each person should have a packet of essays. Some of you are marking essays below pictures; others have no picture but the boys were given one at the time they wrote the story. Some of the stories are two pages long, so watch out for that.

(2) Give each story a ranking on a 0 - 10 point scale. Feel free to use the whole scale. Put the ranking at the bottom of each essay.

(3) An average essay in the package should get five; poorer less; better, more.

(4) Don't use half marks.

(5) You should not judge against normal school children standards, as these essays are by slow learners.

(6) In giving a single overall score, take into account:

i. mechanical aspects, length, spelling, grammar, punctuation

ii. vocabulary, variety and word usage

iii. number of ideas

iv. development of ideas, and

v. creativity.

The subject's mean holistic score per essay was obtained by averaging the independent ratings of the four undergraduate scorers. Then a mean holistic score per base-line and intervention phases per subject was calculated by averaging the subject's mean score for the eight stories selected for evaluation from each phase.

Reliability

Interscorer reliability was essential since scoring the samples of written composition involved varying degrees of subjectivity and skill. Reliability was assessed for both analytic and holistic scoring.

(a) Analytic scoring. Half of the total essay samples were originally marked together by the experimenter and a training assistant (A). The other half were marked by the same experimenter and another trained assistant (B) again using consensus marking. Then 25% of mixed essays in each phase of Experiment 1 were randomly selected per subject, and marked by a third single scorer in order to obtain reliability checks.

Reliability was not obtained for the new words measure because the third single scorer received only a random sample of undated essays and the new words measure depended on a full, dated sequence of essays for its calculation. For all other measures, reliability was calculated by dividing the number of agreements by the number of agreements plus disagreements and multiplying this by 100 (Maloney & Hopkins, 1973). An agreement occurred if both scorers concurred that a response was correct or both said that the response was incorrect. A disagreement occurred when one scorer said a response was correct, while the other said it was incorrect. According to Diederich (1974) a reliability coefficient of 0.80 was high enough for programme evaluation and 0.90 for individual growth measurement in teaching or research.

Table 10 presents a summary of interscorer reliability

Insert Table 10 here

for Experiment 1. Interscorer agreement was generally very high with the lowest mean at 85.53% (total adverbs) and the highest mean at 99.76 (total words written). Of the 27 measures assessed for reliability, the majority (20) received reliabilities in the 90's, with the remaining (7) obtaining reliability agreements between means of 85% and 90%. Ranges varied from a slight difference (95%-100%) to a 75% difference (25%-100%). However they were biased by the small scales used in some measures, for example the thematic maturity score.

(b) Holistic scoring. Each paper was graded by four different scorers using a 0 - 10 scale. Reliability was measured by assessing how many papers scorers agreed on, differed by one point, or a maximum of two points. Differences beyond two points were assessed and considered highly unreliable.

Forty-eight percent of the sample of written compositions selected for holistic scoring were rated reliably. That is, interscorer differences ranged no more than one or two points. On 28% of the stories, scores ranged by three points. 14% of stories received scores ranging by four points; 5% had a range of five points; and 3% received scores ranging by six points. On 2% of the stories the score ranged seven points. That is, the same essay, rated by four different scorers, received scores ranging from three to ten.

Table 10

Summary of interscorer reliability for Experiment 1

Measure	Mean	Range
Handwriting score	94.25	67-100
Total words written	99.76	95-100
Total sentences	99.54	89-100
Total T-units	97.31	67-100
Vocabulary diversity	99.07	94-100
% correct capitalization	95.05 94.98	78-100 80-100
% correct punctuation	94.02 93.12	50-100 50-100
% spelling inaccuracy	88.68	25-100
Vocabulary score	96.29	67-100
Word usage score	86.15	50-100
% mature words	91.76	50-100
grammar errors	95.36	85-100
% total nouns	95.58	67-100
% total A. verbs	89.22 96.41	33-100 57-100
% total adjectives	92.02	50-100
% total adverbs	85.53	25-100
% diff. nouns	96.03	67-100
% diff. A. verbs	88.69 96.41	50-100 50-100
% diff. adjectives	90.58	50-100
% diff. adverbs	86.37	25-100
mean T-unit length	96.81	64-100
Thematic maturity	86.59	67-100
Creativity score	91.12	83-100

Statistical Procedures

Experiment 1 was analyzed using a repeated measures ANOVA. Means for baseline, intervention, transfer and maintenance were entered as the four levels of the single repeated measures factor. Planned comparisons were carried out contrasting the cell means between each successive treatment phase, as well as the difference between baseline and maintenance.

A repeated measures t-test was used to test for differences between baseline and intervention in holistic scores.

RESULTS

Target Behaviour

Figure 2 presents individual data for the measure of

Insert Figure 2 here

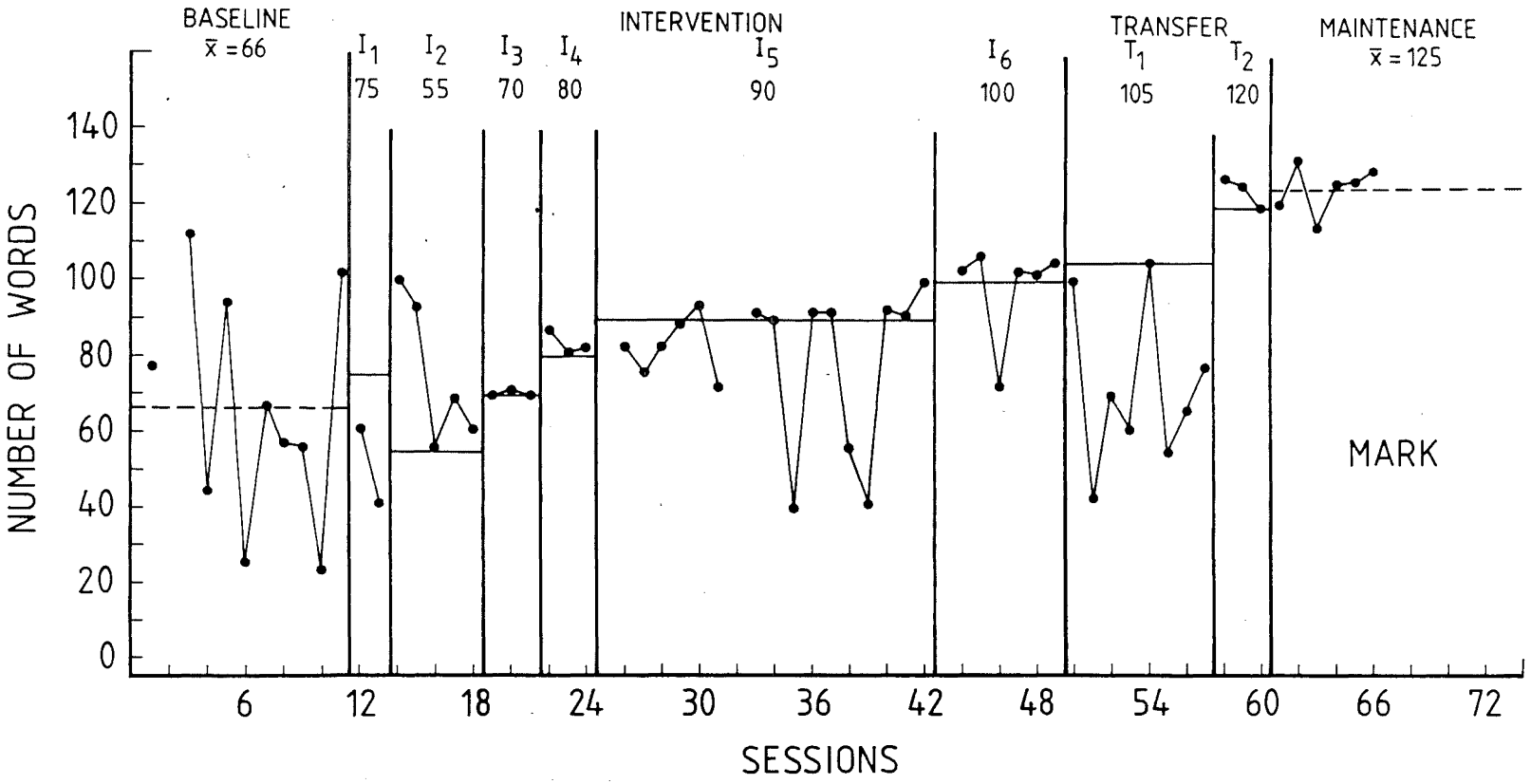
number of words, to which the intervention was directly applied. Writing output was at a low level and variable during baseline. However when intervention was introduced, writing output (Table 11) increased about 50% in the cases of

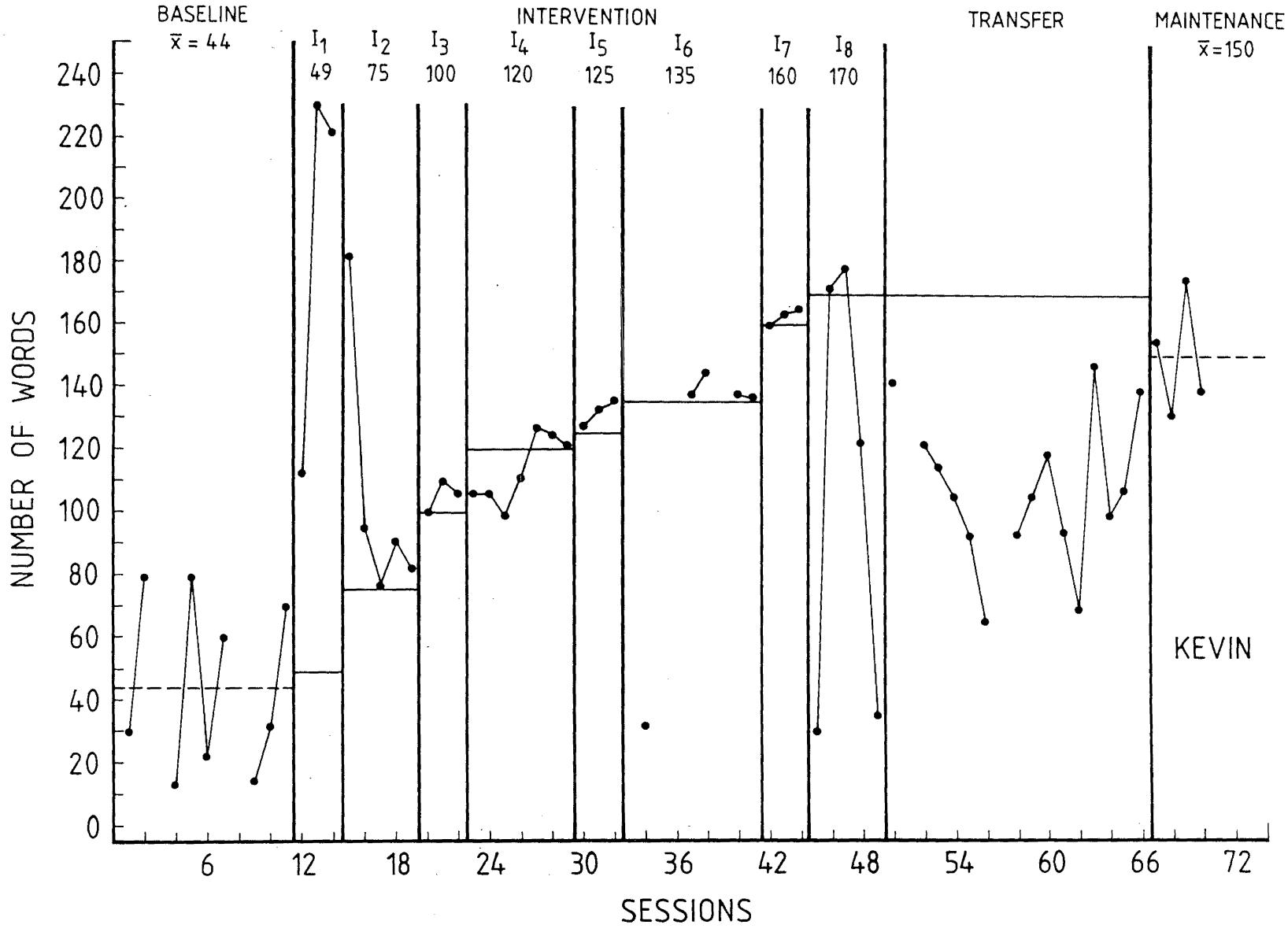
Insert Table 11 here

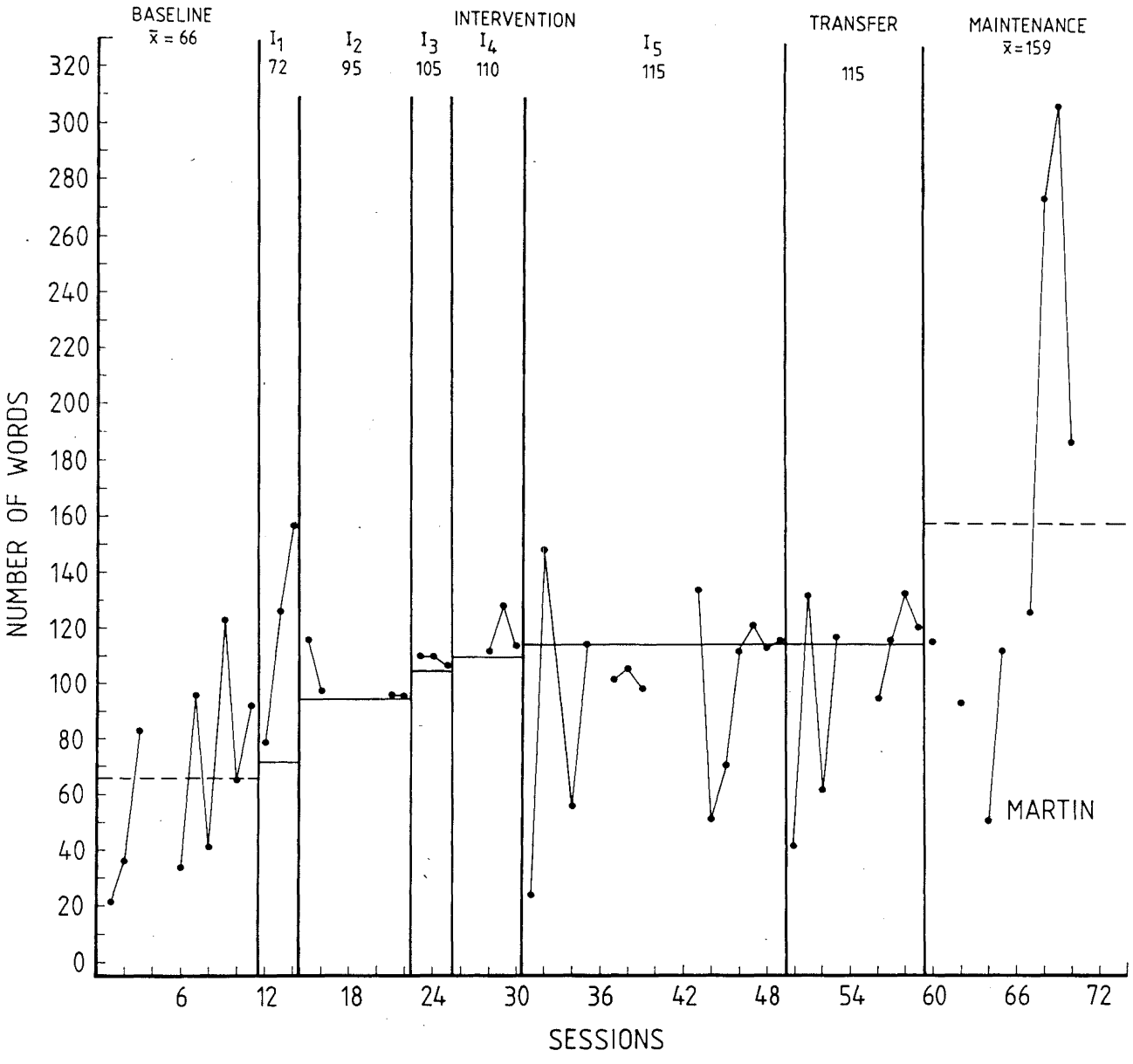
Mark, Martin and Brendon, and just under 200% for Kevin. There was variability in the number of increases that occurred in set levels of written output. For example Kevin set eight different levels, Mark and Brendon set six, and Mark set five different criterion levels. The number of words written remained at the same level during the transfer phase as was

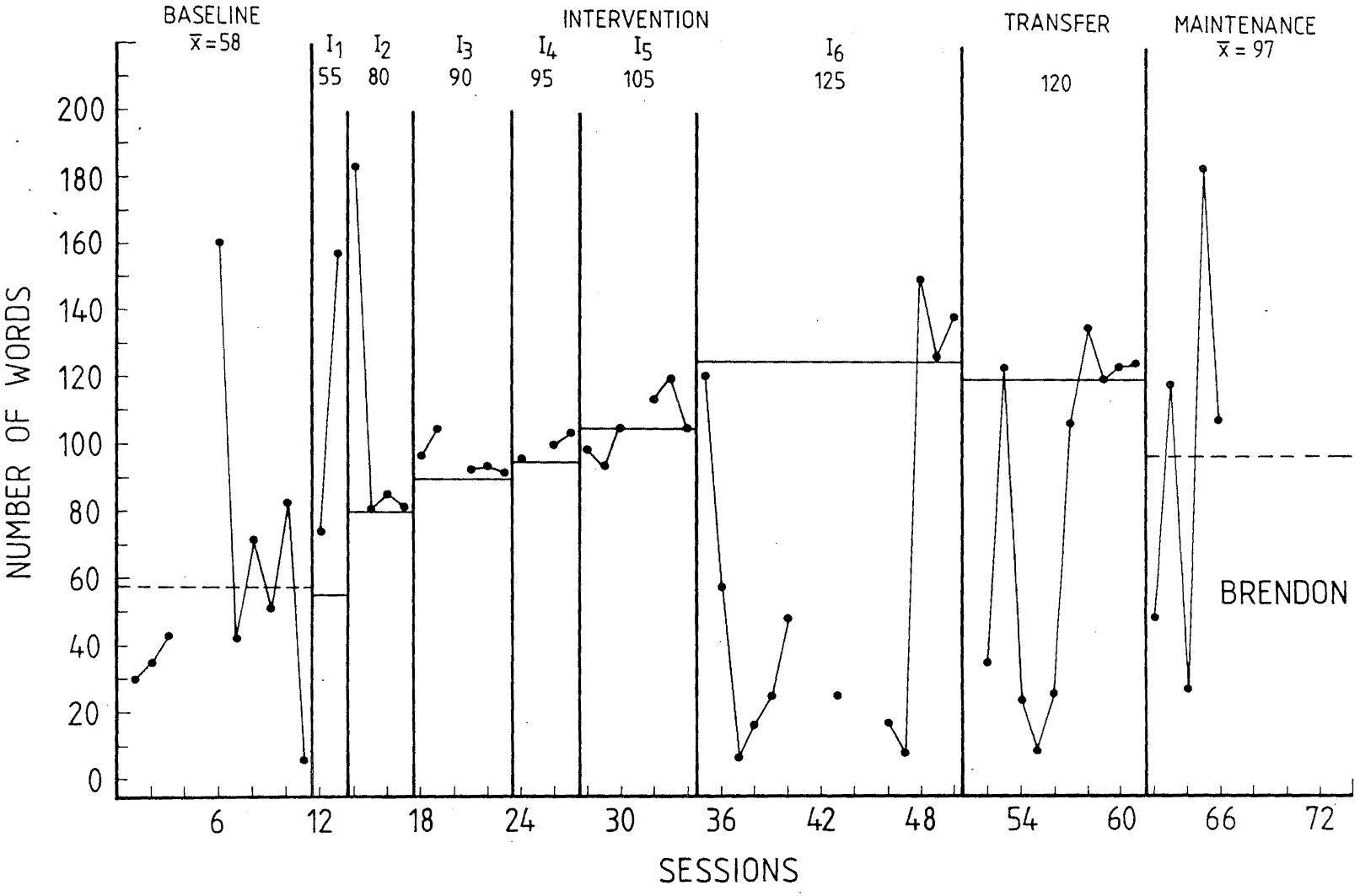
FIGURE CAPTION

FIGURE 2. Number of words for Mark, Kevin, Martin, and Brendon across baseline, intervention, transfer and maintenance phases.


BASELINE
x̄ = 66
INTERVENTION
TRANSFER
MAINTENANCE
x̄ = 125
I1
75
I2
55
I3
70
I4
80
I5
90
I6
100
T1
105
T2
120
MARK
NUMBER OF WORDS
0
20
40
60
80
100
120
140
6
12
18
24
30
36
42
48
54
60
66
72
SESSIONS


BASELINE
x̄ = 44
INTERVENTION
I1 49
I2 75
I3 100
I4 120
I5 125
I6 135
I7 160
I8 170
TRANSFER
MAINTENANCE
x̄ = 150
NUMBER OF WORDS
0
20
40
60
80
100
120
140
160
180
200
220
240
6
12
18
24
30
36
42
48
54
60
66
72
SESSIONS
KEVIN


BASELINE
x̄ = 66
INTERVENTION
I1
72
I2
95
I3
105
I4
110
I5
115
TRANSFER
115
MAINTENANCE
x̄ = 159
NUMBER OF WORDS
0
20
40
60
80
100
120
140
160
180
200
220
240
260
280
300
320
6
12
18
24
30
36
42
48
54
60
66
72
SESSIONS
MARTIN


BASELINE
x̄ = 58
INTERVENTION
I1
55
I2
80
I3
90
I4
95
I5
105
I6
125
TRANSFER
120
MAINTENANCE
x̄ = 97
NUMBER OF WORDS
0
20
40
60
80
100
120
140
160
180
200
6
12
18
24
30
36
42
48
54
60
66
72
SESSIONS
BRENDON

Mean scores across subjects and experimental conditions

Subjects	Baseline	Intervention									Transfer	Maintenance
		I_1	I_2	I_3	I_4	I_5	I_6	I_7	I_8	Mean		
1. Words												
Mark	66	51	76	70	83	80	99			81	87	125
Kevin	44	188	105	105	114	132	118	163	108	124	108	150
Martin	66	121	102	109	118	98				105	103	159
Brendon	58	116	108	96	100	106	62			88	83	97
2. Handwriting												
Mark	4	4	4	4	4	4	5			4	5	4
Kevin											6	7
Brendon	5						5			5	6	6
3. Sentences												
Mark	8	6	7	6	9	8	11			8	8	12
Kevin	4	19	10	12	13	12	11	15	14	13	9	11
Martin	5	10	7	9	10	6				7	5	9
Brendon	5	11	9	8	9	10	5			8	7	9
4. T-units												
Mark	9	6	7	7	10	9	13			9	9	14
Kevin	5	20	11	12	14	12	13	17	15	14	12	16
Martin	7	12	10	14	14	10				11	9	16
Brendon	6	12	11	10	11	11	7			9	8	10
5. New words												
Mark	38	29	18	25	22	18	18			20	14	13
Kevin	47	30	17	13	16	12	9	8	11	14	7	5
Martin	44	23	16	17	15	13				15	14	12
Brendon	48	29	21	16	22	17	18			19	17	12
6. Vocabulary diversity												
Mark	3.7	3.6	3.8	3.8	3.4	4.0	4.2			3.9	3.9	4.6
Kevin	3.0	5.0	4.0	3.9	4.4	4.3	4.6	4.6	4.4	4.4	4.4	4.7
Martin	3.7	4.4	4.3	4.3	4.1	4.2				4.2	4.2	4.6
Brendon	3.3	4.3	3.8	3.7	4.0	4.2	3.2			3.7	3.6	3.5
7. Capitalization												
Mark	73	87	78	54	72	68	83			72	80	76
Kevin	66	72	71	72	75	77	73	80	81	75	77	76
Martin	93	92	94	91	78	86				88	88	86
Brendon	82	84	76	79	64	68	71			72	69	68

reached during the intervention phase. For three boys (Mark, Kevin, Martin) this level increased by 50% during maintenance. The fourth boy increased to a lesser degree.

Statistical analysis showed an overall significant effect for number of words [$F(9,3) = 15.39$, $p < 0.001$] and a comparison between transfer and maintenance showed a significant increase [$F(1,3) = 18.44$, $p < .023$] as did a comparison between baseline and maintenance phases [$F(1,3) = 23.34$, $p < .017$].

Collateral Behaviours

Mechanical component The mean handwriting score for cursive handwriting applied only to Mark. The results in Table 11 show that his cursive handwriting remained of a similar quality throughout the various phases.

Productive component The results for mean number of words have already been discussed. With respect to mean number of sentences, Table 11 shows that three boys (Kevin, Martin and Brendon) increased markedly their number of sentences from baseline to intervention, while the fourth, Mark, maintained the same level of responding. The number of sentences decreased slightly in all but one case (Mark) during transfer of control to the teacher. The level increased again slightly during the maintenance phase.

Statistical analysis showed an overall significant effect for number of sentences [$F(9,3) = 6.18$, $p < .014$] and a comparison between transfer and maintenance showed a significant increase [$F(1,3) = 27.00$, $p < .014$]. The increase from baseline to maintenance was also significant [$F(1,3) = 40.11$, $p < .008$].

The results for mean number of T-units (Table 11) showed a nearly identical pattern to that for sentences. From baseline to maintenance, the increase ranged from approximately 50% (Mark and Brendon) to 100% (Martin) to 200% for Kevin.

Statistical analysis showed an overall significant effect for number of T-units [$F(9,3) = 10.02$, $p < .003$] and a comparison between transfer and maintenance showed a significant increase [$F(1,3) = 18.69$, $p < .023$]. The increase from baseline to maintenance was also significant [$F(1,3) = 19.26$, $p < .022$].

Table 11 presents the mean percentage of new words measure. The highest mean percentage of new words occurred in the baseline phase. In later phases, the mean percentage of new words progressively decreased.

The decrease in the mean percentage of new words between baseline and intervention [$F(1,3) = 71.43$, $p < .003$] and baseline and maintenance [$F(1,3) = 89.48$, $p < .003$] was statistically significant.

The mean vocabulary diversity score (Table 11) increased for all subjects across phases. The diversity score ranged from 3.0 to 3.7 in baseline compared with a range of 3.7 to 4.4 in intervention. The increased level of responding was maintained in the transfer phase and increased again during maintenance, with a range of 3.5 to 4.7.

Conventional component Results for the mean percentage capitalization correct (Table 11) show that although written output increased across phases, this was not associated with an increase in capitalization errors. Levels of responding remained approximately constant from baseline to maintenance with one exception (Brendon) who decreased slightly.

The level of the mean percentage punctuation correct, as seen in Table 12, varied between subjects across phases.

Insert Table 12 here

Brendon made fewer punctuation errors from baseline to maintenance, Mark remained at a relatively consistent level, Martin made progressively more punctuation errors from baseline to maintenance, while Kevin made more errors in transfer and maintenance than in earlier phases. Overall, decreases were not statistically significant.

For two boys (Mark and Kevin) mean percentage spelling errors (Table 12) decreased slightly and for the other two (Martin and Brendon) they increased slightly across phases.

Linguistic component. The first of the three semantic measures, mean vocabulary score (Table 12) increased from baseline to intervention for all boys. Two boys increased moderately (Mark and Brendon) and two markedly (Kevin and Martin). One boy (Martin) showed a further increase during transfer and maintenance phases while the other boys maintained similar levels of responding during intervention.

The mean word usage score rated on a scale of 0-3 (Table 12) increased for all boys from an average of 1 in baseline and intervention phases to 2 in transfer and maintenance phases. Kevin scored 2 consistently through intervention, transfer and maintenance, increasing from a low baseline score of 0. The other boys were more gradual in their improvement.

The increase was statistically significant between baseline and maintenance phases [$F(1,3) = 25.00$, $p < .015$].

Table 12

Mean scores across subjects and experimental conditions

Subjects	Baseline	Intervention								Transfer	Maintenance
		I_1	I_2	I_3	I_4	I_5	I_6	I_7	I_8	Mean		
8. Punctuation												
Mark	76	87	75	87	57	70	75			73	87	76
Kevin	64	81	64	81	66	64	62	67	55	66	52	41
Martin	78	77	75	71	69	66				69	66	53
Brendon	67	87	73	74	79	79	70			75	78	78
9. Spelling												
Mark	6	5	2	5	4	5	4			5	4	4
Kevin	10	4	6	5	6	5	9	4	6	6	7	5
Martin	11	14	14	15	15	12				13	13	16
Brendon	14	9	9	7	13	14	13			11	8	17
10. Vocabulary												
Mark	7	6	6	8	5	10	13			9	8	13
Kevin	4	15	7	9	10	10	11	10	8	10	7	9
Martin	5	16	8	11	8	11				10	12	21
Brendon	7	13	14	13	16	10	7			10	8	9
11. Word usage												
Mark	1	1	1	1	1	1	2			1	2	2
Kevin	0	2	2	2	2	2	2	2	2	2	2	2
Martin	1	1	1	2	1	2				1	2	2
Brendon	1	2	1	1	2	2	1			1	2	2
12. Mature words												
Mark	37	44	32	40	39	41	41			39	36	39
Kevin	32	35	31	32	34	25	27	31	30	31	30	30
Martin	30	32	33	39	37	36				35	37	38
Brendon	39	41	42	39	41	42	42			41	40	38
13. Grammar												
Mark	2	4	1	2	2	2	2			2	3	3
Kevin	6	3	5	1	4	4	3	2	3	4	3	4
Martin	2	4	2	2	2	3				3	2	2
Brendon	2	5	1	4	3	3	2			3	2	1
14. Total nouns												
Mark	26	28	24	35	27	28	26			28	27	26
Kevin	27	23	25	25	21	21	19	18	21	21	18	20
Martin	18	18	21	21	24	23				22	23	27
Brendon	29	24	31	29	25	28	29			28	28	29

The mean percentage level of mature words (Table 12) remained consistent from baseline through to maintenance for Mark, Kevin and Brendon. One exception, Martin, gradually increased from a mean percentage of 30% in baseline to 38% in the maintenance phase.

The first syntax measure was mean grammar errors per 100 words (Table 12). Mark, Martin and Brendon consistently averaged 2-2.5 grammar errors per 100 words across all phases. Kevin made fewer grammar errors from a mean of 6 per 100 words in baseline to 4 per 100 words in intervention, transfer and maintenance phases. None of the differences in this or other syntax measures were statistically significant.

The mean percentage level of total nouns (Table 12) varied between subjects. Mark and Brendon consistently maintained their level of responding across phases. Kevin decreased his mean percentage total nouns from baseline (27%) to maintenance (20%) in contrast to Martin who increased in this measure from baseline (18%) to maintenance (27%).

The results for mean percentage total action verbs (Table 13) varied among subjects across phases. Two boys

Insert Table 13 here

(Martin and Brendon) increased in this measure from baseline to intervention, and increased again in transfer, but then in maintenance decreased to baseline levels. Mark increased his percentage from baseline (39%) to intervention (62%) and although he decreased in transfer and maintenance, the level was still considerably higher than at baseline. Although Kevin did not increase in this measure, he consistently maintained his percentage level of action verbs across phases.

Table 13

Mean scores across subjects and experimental conditions

Subjects	Baseline	Intervention									Transfer	Maintenance
		I_1	I_2	I_3	I_4	I_5	I_6	I_7	I_8	Mean		
15. Total verbs												
Mark	39	69	67	51	74	57	67			62	51	55
Kevin	60	69	60	63	62	67	49	60	50	59	57	61
Martin	65	60	75	56	80	73				71	77	66
Brendon	61	75	77	72	68	78	63			70	73	62
16. Total adjectives												
Mark	9	10	7	7	6	9	5			7	8	8
Kevin	7	8	9	8	7	3	8	7	7	7	7	6
Martin	8	7	6	6	4	9				8	8	8
Brendon	10	9	5	6	9	8	6			7	7	4
17. Total adverbs												
Mark	6	5	7	3	7	8	10			8	8	8
Kevin	6	7	7	9	8	10	8	7	11	8	10	8
Martin	9	7	7	8	7	9				8	8	7
Brendon	5	7	7	3	4	5	7			6	7	7
18. Different nouns												
Mark	22	24	19	25	22	22	19			21	20	19
Kevin	14	18	15	17	15	16	14	10	15	15	13	15
Martin	15	12	14	17	15	17				16	17	19
Brendon	23	15	12	15	19	19	22			18	18	17
19. Different verbs												
Mark	29	58	49	39	55	43	57			48	43	36
Kevin	43	42	43	35	42	36	37	33	31	38	33	41
Martin	55	50	63	44	64	55				55	62	54
Brendon	56	61	53	42	53	58	54			53	60	42
20. Different adjectives												
Mark	8	9	6	7	6	7	5			6	8	7
Kevin	6	7	8	6	6	3	7	5	6	6	5	5
Martin	7	6	6	5	4	8				6	6	6
Brendon	7	6	4	5	7	7	5			6	6	3
21. Different adverbs												
Mark	5	5	5	3	5	7	8			6	6	6
Kevin	4	6	5	6	6	8	7	5	7	6	7	5
Martin	7	6	6	6	5	7				7	6	5
Brendon	6	6	5	2	3	4	7			5	6	5

For three boys (Mark, Kevin and Martin) the level of mean percentage total adjectives (Table 13) was consistent across all phases. However the exception was Brendon whose mean percentage of total adjectives decreased from baseline (10%) to maintenance (4%).

Mean percentage total adverbs (Table 13) increased slightly for Mark, Kevin and Brendon over the various phases and decreased slightly for Martin, from 9% in baseline to 7% during maintenance.

For two boys (Mark and Kevin) the mean percentage different nouns (Table 13) was consistent across phases. For Martin the percentage levels increased from baseline (15%) to maintenance (19%). This contrasted with Brendon where his level decreased from baseline (23%) to maintenance (17%).

The results for mean percentage different action verbs (Table 13) varied between subjects and across phases. Mark increased greatly from baseline (29%) to intervention (48%) however this level decreased in transfer and maintenance, but was still higher than baseline levels. Kevin had similar percentage levels in baseline and maintenance in contrast to lower percentage levels in intervention and transfer phases. Martin maintained a consistent level of responding across all phases bar the transfer phase where the level increased moderately. Brendon's results were consistent across baseline and intervention, increased in the transfer phase but then decreased in the maintenance phase.

With regard to mean percentage different adjectives (Table 13) all boys responded at a similar level across all phases, bar Brendon who decreased 50% from transfer to maintenance.

All four subjects maintained a nearly constant percentage level of responding across all phases with respect to mean percentage different adverbs (Table 13).

The figures in Table 14 show the percentage decrease

Insert Table 14 here

when mean percentage different parts of speech were subtracted from mean percentage total parts of speech. This subtraction provided a measure of repetition of parts of speech within essays. The mean percentage decrease for nouns was 7% (range 3-13); for action verbs 15% (range 5-24); for adjectives 1% (range 0-3); and for adverbs 2% (range 1-3).

The mean T-unit length (Table 15) was similar in base-

Insert Table 15 here

line and intervention for all boys, increased slightly between intervention and transfer and this level of responding remained in the maintenance phase.

Creativity component The mean thematic maturity score (scale 1-10) increased from a mean of 4 in baseline for all boys to a mean of 5 in intervention. This level of responding was maintained in transfer and maintenance phases as shown in Table 15.

The mean creativity score (Table 15) increased greatly for all boys from baseline to intervention. This level of responding remained similar in transfer and maintenance phases for Mark, Kevin and Brendon. However the level increased for Martin during transfer and maintenance phases.

Statistical analysis showed an overall significant effect for the creativity score [$F(9,3) = 11.45$, $p < .002$] and a comparison between baseline and intervention demonstrated a

Table 14

The decrease in percentage from total to different parts of speech.

Subjects	Phases			
	Baseline	Intervention	Transfer	Maintenance
1. Nouns				
Mark	4	7	7	7
Kevin	13	6	5	5
Martin	3	6	6	8
Brendon	6	10	10	12
2. Action verbs				
Mark	10	14	8	19
Kevin	17	21	24	20
Martin	10	16	15	12
Brendon	5	17	13	20
3. Adjectives				
Mark	1	1	0	1
Kevin	1	1	2	1
Martin	1	2	2	2
Brendon	3	1	1	1
4. Adverbs				
Mark	1	2	2	2
Kevin	2	2	3	3
Martin	2	1	2	2
Brendon	1	1	1	2

Table 15

Mean scores across subjects and experimental conditions

Subjects	Baseline	Intervention									Transfer	Maintenance
		I_1	I_2	I_3	I_4	I_5	I_6	I_7	I_8	Mean		
22. T-unit length												
Mark	8	10	11	10	8	10	8			8	10	10
Kevin	10	10	10	9	8	11	10	9	7	9	9	10
Martin	10	10	11	8	9	10				10	11	10
Brendon	9	10	10	10	9	11	10			10	11	10
23. Creativity												
Mark	10	12	12	12	15	14	18			14	15	16
Kevin	8	15	14	15	17	16	17	15	14	15	16	16
Martin	10	15	13	13	14	18				16	21	27
Brendon	10	18	19	19	19	17	13			16	16	17
24. Thematic												
Mark	4	4	4	4	5	5	5			5	5	5
Kevin	4	5	5	5	6	4	5	6	3	5	5	4
Martin	4	5	4	6	5	5				5	5	6
Brendon	4	6	5	6	5	5	3			5	5	5
25. Holistic score												
Mark	4									5		
Kevin	3									5		
Martin	4									6		
Brendon	4									3		

significant increase [$F(1,3) = 83.53$, $p < .003$]. The increase between baseline and maintenance phases was also significant [$F(1,3) = 14.06$, $p < .033$].

Holistic Scoring

As shown in Table 15 the essays of three boys (Mark, Kevin and Martin) rated on a scale of 1-10 qualitatively improved from baseline to intervention. The exception was Brendon whose mean holistic score decreased from 4 in baseline to 3 in intervention.

DISCUSSION

This study focussed on written composition and, in particular, the productivity aspect of this broad academic behaviour. Interventions successfully applied to productivity have included mainly consequential procedures such as token reinforcement and instructions (Brigham et al. 1972; Glendinning, 1977) and oral and written comments (McKessar, 1977), oral instructions and praise (Kraetsch, 1981) and performance feedback packages (Scriven & Glynn, 1983; Van Houten et al., 1974, 1975; Van Houten & McKillop, 1977). Researchers have largely ignored antecedent stimuli.

Experiment 1 used the correspondence training procedure (Risley & Hart, 1968) which focussed on both antecedent and consequent stimuli. Although the correspondence training paradigm had been used with children and adults, it had rarely been used to change children's story writing behaviour. The two studies that had applied correspondence training to story writing behaviours (Christie & Ballard, 1983; Mandler & Monsen,

1985) did not focus directly on word output, which was the target behaviour in this study. This experiment was not only unique in applying correspondence training to word productivity, it was also the first study to use correspondence training on a high school age group of slow learners, using a changing-criterion design (Hartmann & Hall, 1976) and monitoring 23 collateral measures.

Experiment 1 aimed to answer five experimental questions which are now discussed.

1. Did correspondence training increase the target behaviour?

The results clearly showed that correspondence training was effective in increasing word productivity. In contrast to the low and variable mean number of words written during baseline, during the intervention phases when correspondence training was implemented, written output increased approximately 50% for three subjects and just under 200% for the fourth. This increase could be attributed to the correspondence training procedure rather than to history, maturation or measurement factors because the performance of each subject matched successive criteria fairly closely.

Another advantage of correspondence training was that subjects could work at their own pace. For example, during the intervention phase one subject set eight target levels, two subjects set six, and the last subject set five.

2. Did correspondence training generalize to collateral behaviours?

The results suggested that the correspondence training procedure, applied solely to the target behaviour of mean number of words written, was effective not only in increasing

that target behaviour, but also in generalizing to either maintain or increase levels of responding in the majority of collateral behaviours.

Correspondence training was effective in generalizing improvement from baseline to intervention in measures mainly from the productive and creative components, namely: mean number of sentences, T-units and percentage total action verbs (three subjects increased, the fourth remained at the same level); and increasing for all subjects the mean vocabulary diversity score, the mean vocabulary score, the mean thematic maturity score (10% increase). The increase for the mean creativity score was statistically significant. These improvements occurred despite correspondence training focusing solely on story length and not on essay quality.

Perhaps the most important of these increases was in the vocabulary diversity score which increased for all subjects from baseline to intervention. This finding suggested that the increase in the mean number of words written from baseline to intervention was not achieved through word repetition. McKessar (1977) also found this result. Additional support for the finding that word output was not based on repetition came from the comparison between total (including repeats) and different (excluding repeats) parts of speech. When the mean percentage "different" parts of speech were subtracted from the mean percentage "total" parts of speech, the percentage decrease for nouns, adjectives and adverbs was only slight. These low decreases demonstrated that the majority of words in the mean percentage total parts of speech measures were different and nonrepetitious. Action verbs, the fourth part of speech, decreased the most, though not greatly,

but this would be expected for this part of speech. For example in an essay on a skiing trip, one would expect the verb "to ski" to be repeated.

The increases in collateral behaviours for mean number of sentences and T-units would follow from the increase in the target behaviour (mean number of words). However the significant increase in the mean creativity score is a genuine generalized increase resulting from correspondence training.

Although increases were not obtained on some of the measures, they did remain at the same level from baseline to intervention. This suggested that reinforcing word output did not lead to deterioration in accuracy or quality of these measures. The latter fell within the mechanical, conventional and linguistic components and included the following measures: T-unit length; mean handwriting score (applicable to three subjects); mean percentage spelling errors and total adjectives (two subjects remained at the same level and two decreased slightly on both measures); mean percentage total adverbs (two remained at the same level and two increased slightly); mean percentage punctuation correct, total nouns and different action verbs (two remained at the same level, one increased slightly and one decreased slightly); mean percentage capitalization correct (one remained at the same level, one increased and two decreased slightly); mean word usage score (three remained at the same level, and one increased greatly); mean percentage mature words and different adverbs (three remained at the same level and one increased slightly); and mean percentage different adjectives and nouns and mean grammar errors per 100 words (three remained at the same level and one decreased slightly).

The lack of increase in the T-unit length measure was not surprising since this index increases gradually with age and grade. Cooper (1975a) suggested that between 0.25 to 0.50 words per T-unit per year was the expected growth, and Experiment 1 was only conducted for approximately five months (including transfer and maintenance phases). In addition changes in T-unit length may not have been detected because of the small total written output. O'Hare (1973) recommended that a sample size of approximately 400 words was necessary for a reliable calculation of mean T-unit length. The consistent results for mean T-unit length indicated, if nothing else, that topics were adequately homogeneous in difficulty, as fluctuations in mean T-unit length have been known to occur because of variation in topic difficulty and discourse type.

The only decrease that occurred was in the mean percentage new words measure which decreased significantly from baseline to intervention. The decrease was also found in Brigham et al's (1972) study. The fact that new words still occurred at a moderate rate through the various experimental phases indicated that the stories were still diverse in content and not merely repetitious. It also showed that the picture prompts were dissimilar enough to stimulate new words throughout the various phases.

It was important to assess collateral behaviours to determine whether the correspondence training procedure increased the target behaviour at the cost of other behaviours. Overall there was deterioration in only one measure, which was expected; otherwise, there was improvement in productive and creative measures; mechanical, conventional and linguistic measures remained at the same level of responding. These

results were achieved despite being neither programmed for nor attended to.

3. Was there a subjective improvement?

Subjective holistic ratings of a sample of written compositions from baseline and intervention phases showed that in three subjects there was a statistically nonsignificant improvement in quality. This improvement corresponded to the analytic scoring results. Other studies (Brigham et al., 1972; Glendinning, 1977; McKessar, 1977; Van Houten et al., 1974 1975; Van Houten & McKillop, 1977) have also demonstrated that contingencies on written output have resulted in essays being judged as more creative or qualitatively improved. The only two studies which applied correspondence training to story writing behaviours, though not to number of words written, found a minor improvement in rated story quality (Mander & Monsen, 1985) and no improvement in the subjective quality data between baseline and intervention phases (Christie & Ballard, 1983).

Unlike the three subjects who improved slightly, the fourth subject decreased slightly in holistic ratings from baseline to intervention. This decrease can be explained in terms of the subject having been ill during the final three weeks of the intervention phase. He wrote only intermittently and therefore the sample of his intervention essays (I_6) chosen for holistic scoring may not have been a valid representation of his true ability. Thus the biased sample, confounded by his illness at that point may explain why both holistic (and analytic scoring) results for this intervention phase were lower than baseline levels.

4. Were improvements sustained during the transfer phase?

The fourth experimental question posed was to assess if correspondence training could be transferred from the experimenter's control to the classroom teacher, without loss of intervention gains. This question was particularly important as correspondence training is an intervention which ideally should be implemented by the teacher in the classroom setting.

With regard to the target behaviour, the mean number of words remained at the same level between intervention and transfer for two subjects, increased for the third and decreased for the fourth. Nevertheless, the mean number of words still exceeded baseline levels for all subjects.

With respect to measures of collateral behaviours, the only decrease was for the new words measure and the reasons for this drop have already been explained. Three collateral behaviours increased from intervention to transfer phases, the mean handwriting score (applicable only to two subjects), the mean word usage score, and mean T-unit length, which slightly increased for three subjects, the fourth remained at the same level.

Most measures of collateral behaviours remained at the same level of responding as during the intervention phase. This occurred for productive measures (mean vocabulary diversity score; and mean number of sentences and T-units - two subjects remained at the same level and two decreased slightly); conventional measures (mean percentage capitalization correct - three remained at the same level and one subject slightly increased; mean percentage punctuation correct - two remained at the same level and one increased and one decreased slightly; and mean percentage spelling errors - three remained at the same level and

one decreased); linguistic measures (vocabulary score, mature words; grammar errors; mean percentage total and different nouns, adjectives and adverbs; mean percentage total action verbs - two remained at the same level, one increased and one decreased slightly; and mean percentage different action verbs - two subjects increased and two decreased slightly); and creative measures (thematic maturity; creativity score - three subjects remained at the same level and one subject increased). Although some boys decreased slightly, this was balanced by other boys increasing, and slight decreases tended in most cases still to be equal to or above baseline levels.

These findings suggested that the classroom teacher's initial attempts at implementing the correspondence training procedure were satisfactory. Although there were not many improvements, gains made during interventions were not lost. As found by Mander and Monsen (1985), the classroom teacher stated that correspondence training was not difficult to implement. In fact she stated that she felt confident to implement it next year and suggested she would supplement it by offering students access to vocabulary sources. Because correspondence training involved few personnel and was easy to apply, it was a practical and inexpensive procedure for the classroom.

5. Were target levels sustained in the maintenance phase?

Although the maintenance phase occurred for only approximately ten writing sessions (because the school term was concluding) the target behaviour (mean number of words) increased significantly between transfer and maintenance and baseline and maintenance phases. Even though the correspondence

between the verbal statement (telling the teacher the number of words intended to be written) and the performance (the total number of words actually written) was no longer reinforced, the level of responding on collateral behaviours either increased (especially productive measures) or remained the same as the level in the transfer phase. This demonstrated that the subject could write a qualitatively and quantitatively acceptable story when only his statement of intended word output was reinforced. The results suggested that "doing" was under the stimulus control of "saying" as high levels of behaviour resulted merely from reinforcing "saying".

Thus it can be concluded that correspondence training is an effective alternative to traditional methods of modifying writing behaviour, and may possibly be more valuable as it requires less time to implement and maintain levels of responding.

EXPERIMENT 2

INTRODUCTION

Creative aspects of written composition

Most of the focus in written composition research has been on mechanical correctness. This was due to the traditional esteem in which elements of writing form were held (Connors, 1985). Creative ideas, when considered, were seen as problematic because of issues such as subjectivity, evaluation and standards of what good ideas were. This anti-creativity theme was illustrated by Evans' (1967) statement that the best way to teach the usage of language was to make the child write about tangible, permanent things rather than express feelings. Evans (1967) suggested that a mortal disservice was done to the child when self expression, which he described as mere froth and self-indulgance, was permitted. Evans (1967) believed in disciplined education, with the aim to eliminate errors.

His approach was in direct conflict with researchers and educationalists of the 1970s and 1980s who emphasized reinforcement for the production of ideas (content), originality, or creativity. Once content was achieved, form could be improved (Bennett, 1983; Glynn, 1981; Lickteig, 1981; Pearce, 1983; Simms, 1983; Tripp, 1978). Hence there was a distinction between good English and correct English (Strickland, 1960). Cooper and Odell (1978) have suggested that discovering what the child wanted to say was the basic problem in writing, not decisions about how the ideas were best to be presented.

Emphasis on mechanical correctness has been downplayed somewhat in response to consistent research findings that the correlation between grammatical knowledge and becoming a better writer is very weak (Braddock et al., 1963; Dagenais & Beadle, 1984; Jones, 1972; Koch, 1982; Kuykendall, 1975; Strickland, 1960; West, 1967; Wilkinson, 1964). O'Dea (1965) has suggested that this assumption was one of the most durable myths in written composition. Sherwin (1969) concluded after a review of the past 50 years' research that instruction in the mechanics of grammar was both inefficient and ineffective in achieving writing proficiency. In fact it has been suggested that grammar and mechanics instruction may have a negative effect on writing (Braddock et al., 1963; Moffet, 1968). Carlson (1973) suggested that emphasis on mechanics to the exclusion of content and the child's original thoughts would elicit imitative, woodenly worded composition or lead to a complete rejection of writing.

Kinnick (1960) suggested giving two grades, one for the student's effectiveness in achieving his goal and the second grade for mechanics. However Judy (1973) argued that double grades merely avoided the form/content issue. Evertts (1966) suggested that if ideas were emphasized, then content not mechanics should be evaluated. Halvorson (1960, cited in Braddock, 1969) found that sixth graders learned whatever was emphasized, whether mechanics or ideas. If this was the case, and emphasis had been on mechanics, then it is not surprising that Glynn (1981) suggested that certain aspects in school writing programms were being extinguished because of the predominant reinforcement of spelling, handwriting, accuracy of form — in short, mechanics. According to Bording et al. (1984) it is more effective to teach grammar skills through

creative writing than in isolation, since transfer of grammar skills to creative writing is questionable. Poplin et al. (1980) concluded that immediate, appropriate feedback to writing offers the most effective means of improving the mechanics of writing without stifling attention to content.

Freedman's (1979) work has shown that teachers respond to mechanics and sentence structure only when the development and organization of ideas is under control. However Harris (1977) found that despite stating that content and organization were more important, teachers were influenced by mechanical errors even when the essay had good content and organisation. Marshall (1971) found that comments made by high school English teachers on student compositions focussed on technical errors and were restricted mainly to symbols and abbreviations. In Kline's (1976) list of college composition teacher's marking priorities, the focus was overwhelmingly on form and errors. With respect to the intermediate grade level, Searle and Dillon (1980), who investigated teacher written responses to composition, confirmed this general trend. Their results suggested that teachers responded to form rather than content. In addition the written responses tended to be a simple general comment or a mark, and a correction of all mechanical errors. Therefore, contrary to research advice, it appeared that teachers were still preoccupied with form (Martin, 1975; Pearce, 1983).

What is creativity?

Any article on creativity will remind the reader that the elusive concept of creativity is difficult to define. It may be generally defined as the process of recombining unusual

elements or thoughts into new, valuable and meaningful relationships. Others (e.g., Kagan, 1967; Maltzman, 1960) have regarded it as a product of original behaviour and traits such as flexibility, fluency, motivation and temperament. Parnes (1967) alternatively defined creative behaviour in behaviouristic terms as a response or pattern of responses which operate upon discriminative stimuli (things, words or symbols) to result in at least one unique combination that reinforces the response. He classified such creative behaviour as discriminative, manipulative and evaluative.

Lloyd-Jones (1970) stressed the importance to teachers of having a specific and precise definition of creativity. Another confounding factor is that creativity judgement is usually dependent on the norms of a given society or the subject's past behaviour (Maloney & Hopkins, 1973; Maltzman, 1960).

These issues have resulted in ambivalence toward the experimental study of creativity. For example Holman, Goetz and Baer (1977) suggested that the subjectivity of creativity leads to tentative analyses being rejected as having missed the point or the "true" behaviour. Gutman (1961) stated that creative behaviour is by nature spontaneous, inner-directed, and ordinarily not capable of being elicited at will. Therefore, it is difficult to manipulate and control, and not amenable to experimentation. Likewise Davies (1969) suggested the ambiguity of the term creative writing makes it impossible to evaluate the originality, sensitivity or imaginative content of a piece of writing. Nevertheless research has been attempted and encouraged (Holman et al., 1977). For example, Moslemi (1975) concluded from a review of the literature that

future research needs to focus on defining and grading creative writing.

Rosenbaum (1968) used the terms novelty and reinforcement to distinguish between creative and non-creative writing. Creative writing he suggested is high in uncommon responses, whereas non-creative writing is high in expected responses. Not only must the novelty be statistically infrequent, it must also receive audience response or reinforcement. The notion of audience validation determining whether or not the writing is creative, was held important in this thesis and will be elaborated on in a later section. This thesis adopted the definition of creative writing provided by Moslemi (1975) as a "broad general term describing the characteristics of a student's composition in respect to its originality, idea production, language usage and uniqueness of style". Originality was defined as use of new, imaginative or unusual ideas, or a common idea used in a new and imaginative fashion. It was an uncommon response to a stimulus. Idea Production referred to the quality, quantity, fluency or diversity of ideas or precise, detailed description or elaboration of one person, object, experience, or idea. Language usage was defined as the use of imagery, lively description and figures of speech (metaphor, simile, personification, etc.) and the coining of new words. Finally, uniqueness of style reflected the writer's unique individuality, particular preferences, moods, tastes, beliefs, humour or wit.

The area of creativity became the object of research during the 1950s, perhaps in response to Guilford's (1950) paper in which he pointed out and gave reasons for education's appalling neglect of the study of creativity. For example,

only 186 of the 121,000 titles indexed in psychological abstracts focussed on creativity. Guilford hypothesized not one but several factors or primary abilities as being related to creative performance: sensitivity to problems; synthesizing and analyzing abilities; a reorganization or redefinition ability; span of conceptual or ideation structure; an evaluation ability; ideational fluency, flexibility of set, and ideational novelty. The latter three factors have kept the attention of researchers over the past three decades. Fluency referred to the number of ideas or solutions a person was capable of producing. A creative person would generate a large number of ideas in his/her area of creativity. Flexibility emphasized the number of different kinds or categories of ideas (Shallcroft, 1981). The creative person would generate ideas in categories other than the common ones. Novelty or originality of ideas referred to the degree of uniqueness of ideas as compared with the ideas of others in the group (Shallcroft, 1981). The ideas were usually statistically uncommon. A fourth factor referred to was elaboration. This reflected the number of ideational details which were worked out. These factors could apply to both product and process. A more indepth discussion of these factors can be found in Gowan, Demos and Torrance (1967) and Torrance (1967).

Because Experiment 2 dealt with creativity in mildly retarded children, the common assumption that low intelligence groups are not capable of creative ideas was examined. A review of the literature on the relationship between creativity and intelligence failed to support this assumption (Ford & Renzulli, 1976; Getzels & Jackson, 1962; Lloyd-Jones,

1970). Although a minimum of intelligence was a necessary prerequisite, creativity had a weak correlation with IQ scores. This finding was not surprising when one considered that an IQ test may not be a valid indicator of the individual's true performance. It was disproportionately loaded on the verbal comprehension factor and had low face validity with regard to creative content. Considering that an IQ test does not tap intelligence per se, a weak correlation was not an unexpected outcome (Barron, 1969).

Creative aspects of written composition: Behavioural intervention

Perhaps the first experimental studies applying behavioural principles to creativity were a series by Maltzman (1960) which showed that verbal instruction and/or praise increased original word associations in college students. He focussed on the originality factor of creativity which had both face validity and was more open to intervention and objective measurement. Goetz and Baer (1971), however, suggested that creativity is most meaningful when it is a description of the individual's behaviour rather than expressed as changes in the means of different groups.

Torrance (1965) found that reward contingent on originality resulted in a high level of interest and originality with sixth graders. Reward contingent on correctness resulted in gains in that domain alone. However the study had few details on the subjects, omitted a control group, did not give intermarker reliability coefficients, was based on one story sample per subject, and did not control title stimulus prompts. The originality inventory had not been validated and the

correctness scale was vauge and too global.

In the 1970s researchers attempted to increase new form production and form diversity in more relevant behaviours, namely blockbuilding (Goetz & Baer, 1971, 1973), easel painting (Goetz & Salmonson, 1972; Holman et al., 1977), tools (Parsonson & Baer, 1978), felt-pen drawing (Holman et al., 1977) and doughmodelling and storytelling (Trenberth & Parsonson, 1985).

Research in the 1970s also focussed on the more complex behaviour of written composition. Brigham et al. (1972) applied sequential token reinforcement contingencies and minimal instruction to number of words, number of different words and number of new words to the writing of 5th graders with academic and behaviour problems. While this resulted in increases in the length of the stories, there was little effect on redundancy and vocabulary. Overall ratings of the stories for mechanics, vocabulary, ideas and internal consistency were highest when new words were the target of reinforcement, with number of words being second, and there was little effect on different words. A subjective change was also found with an increase in interests and attitudes of the boys to written composition. Unfortunately, the study did not present reliability ratings of judges, subjects were not described in much detail and both length of writing session and amount of reinforcement were insufficiently controlled across groups. In addition, maintenance of results was neither programmed nor monitored.

Lovitt (1976) selected six elements covering both content and mechanical aspects of creative writing (word frequency, punctuation and capitalization, mean sentence

length, number of different words and a quality measure). He found that feedback on mechanics led to improvements to both mechanics and content. However, feedback on content may not have been as effective as feedback on mechanics because content feedback involved negative statements which may have had a more deleterious effect on an individual's story content than mechanics. Lovitt (1976) also viewed content of creative writing as consisting of style, syntax and description, which excluded a substantial amount. Finally, ten minutes seems too short to write a story, especially with learning disabled students.

Glover and Gary (1976) demonstrated that instructions, team points, and practice could increase number of different responses (fluency), number of verb forms (flexibility), number of words per responses (elaboration) and statistical infrequency of response forms (originality), as well as scores on the Torrance tests of creativity in fourth and fifth graders. However, the individual effects of the three components of the intervention cannot be assessed from the design. Glover (1979) later applied team points and practice to fifth graders' storywriting. Fluency, flexibility and originality, assessed through five measures, all increased as did creativity test scores.

Campbell and Willis (1978) made social and token reinforcement contingent on elaboration in the first phase, included flexibility in the second phase, and fluency in the third. There was an increase in Torrance Test of Creative Thinking scores. In addition, the reinforcement procedures both increased and maintained, for 12 days, elaboration, flexibility and fluency.

Harrop and McCann (1984) replicated the Campbell and Willis (1978) study with third year pupils and found increases in flexibility, fluency and especially elaboration. The intervention consisted of a promised letter to parents based on long term improvement, minor instructions and points accompanied by teacher comments. Interscorer reliability was high considering that teachers marked essays according to their own criteria of creativity. A final phase one month after the study showed maintenance of results. The correlation between teacher judgements and increased scores suggested valid increases in creative writing performance.

Jerram's (1985) study, unlike the majority of studies which applied external teacher contingencies, examined the effect of noncorrective feedback (personal written comments by teacher and later parents) related only to content and ideas. The normal grade five class (mean age ten) wrote for 15 minutes, four days weekly inbetween spelling and mathematics. Intervention led to increases in quantity (writing rate), quality (based on a rating scale and advanced words used) and spelling accuracy. However the latter result may have been confounded by the spelling lesson which occurred consistently prior to the writing time. Another confounding variable could have been the oral feedback on content and ideas given during baseline, which was elaborated on in intervention with similar, but written, comment. When content feedback was written by parents, educators observed an improvement in quality of writing.

<u>The relationship between subjective creative ratings and action verbs</u>

A consistent finding across several applied behavioural studies has been that the highest rating of creativity occurs when the contingency is on action verbs.

Maloney and Hopkins (1973) evaluated the effects of certain procedures (giving examples before each session, a written statement of the contingency on the blackboard, team competitions and token system) contingent on the use of different adjectives, different action verbs and different sentence beginnings in the ten-sentence stories of fourth to sixth graders enrolled in a non-remedial summer school. These variables increased and led to increased subjective ratings of creativity, especially when different action verbs were being manipulated. The only effects on collateral behaviours were for adjectives, a response class similar to verbs. Like former studies they defined creativity along the lines of response diversity, in terms of different or nonrepetitive responses. Unfortunately response definitions were not greatly detailed, nor was the experiment conducted for a reasonable length of time (e.g., four days for each intervention), and it was unclear which combination of independent variables affected the dependent variables. Ordering the stories on a 1 to 4 scale of creativity told little about what criteria were used, and whether in fact they were valid and consistent both within and across the raters. This argument was supported by the reliability between independent raters which was only 46%.

Two years later, Maloney, Jacobson and Hopkins (1975) evaluated the effects of teacher lectures, requests, praise and free time, contingent on different parts of speech

(action verbs, adjectives, adverbs, and both the latter parts of speech) on the five-sentence stories of third graders. Interestingly, lectures and requests to use particular parts of speech were not successful in changing writing behaviour. The addition of positive reinforcement was necessary for improvements to occur.

Like Maloney and Hopkins (1973), Maloney et al. (1975) also found that creativity judgements were highest for stories written when contingencies were placed on action verbs. Effects on collateral behaviours were also found, with a slight increase in total number and number of different adverbs when action verbs were manipulated, and a greater increase in action verbs when adverbs were manipulated. They suggested that attention in future research should be focussed on developing strategies both to maintain and reinforce creative behaviour.

Instead of examining the teacher's execution of strategies, Ballard and Glynn (1975) assessed pupil self-assessment and self-recording procedures, with partially self-determined and self-administered reinforcement contingencies (token reinforcement) on number of sentences, different action words and different describing words (including both adjectives and adverbs). Results showed that the total self-management procedures could be effectively carried out. Writing responses increased and their quality improved. However self-assessment and self-recording of responses alone had little effect on writing. As in Maloney et al's (1973, 1975) study the highest quality ratings occurred when different action words were self-reinforced. However, order effects were not controlled for in the design,

Additionally, there was an increase in on-task behaviour in the self-reinforcement phases. Subject details and the action word response definition were not very elaborate; however intervention phases were of a more reasonable length than in Maloney and Hopkin's study (1973). As with the latter study, a 1 to 5 scale was used, which lacked descriptive results and its holistic nature seems open to interpretation.

The results of Ballard and Glynn's (1975) study lead one to ask if self-management procedures could be a viable option with special education students. In fact this issue was examined by Glynn, Wotherspoon and Harbridge (1976) who described the long term effects of a token reinforcement study (Glynn, McNaughton & Wotherspoon, 1974; Wotherspoon, 1974) on attending (group contingency) and writing behaviours (individual contingencies) in a primary school age special class (IQ range 61-72). Intervention resulted in increases in both attending behaviours and generating one's own sentences and words, with mixed results in the writing measures of copying and transcribing words. These increases were maintained in follow-up several months later. At the time of Glynn et al's (1976) study, some of these subjects had moved to a senior special class. They had maintained their improvement in number of words and sentences gained during the original treatment and follow-up. Glynn et al. (1976) increased this sample and assessed whether Ballard and Glynn's (1975) self-management procedures with normal children could succeed with special children. The target behaviour was different action words and intervention resulted in varying (due to confounding variables) increases in mean

number of different action words per day and phase. Some subjects did not cope with the self-monitoring condition, but most succeeded with the change to external reinforcement. The authors suggested that a large amount of training was necessary to enable target behaviours to be recognized reliably by the subject.

Glendinning (1977) investigated the effects of both token reinforcement and instructions on both mathematics and writing (number of words, action words, describing words) and attending behaviours. The subjects had a mean age of about eight years with some academic retardation and social maladjustment. Reinforcement of academic behaviour led to increases in both that behaviour and attending behaviour, in contrast to dual reinforcement which led to lesser increases in both behaviours; and reinforcement of attending behaviour alone which led only to increases in that behaviour with no effects on academic behaviours. Although qualitative ratings were reliable, they were still based on a crude 1 to 5 scale and were based on a small sample (the last story of each condition). Those who received the highest subjective ratings had a higher action to describing word ratio.

Following on from former studies on the performance feedback system (Van Houten et al., 1974, 1975; Van Houten & Van Houten, 1977). Van Houten (1979) examined whether the results of the intervention would generalize to the same behaviour at a different time of the day. The treatment package of explicit timing, self-recording of number of words and public posting was sequentially applied to words per minute and number of different action words. The procedure led to an increased word rate, which generalized to a second

story later in the day. However, with respect to number of different action words, similar results occurred with only the grade 3-4, not the grade 2-3 class. Unlike previous researchers, Van Houten (1979) found no increase in the subjective ratings of story quality with the grade 3-4 class whose different action word total increased with intervention. Van Houten (1979) attributed generalization and maintenance at one- and three-month post experiment checks to the student comments the intervention produced. However, it is questionable whether the seven minutes allowed for writing in Van Houten's (1979) study was long enough for the product to be defined as a story.

This review has demonstrated that, with the exception of Van Houten (1979), subjective ratings of quality were highest when the contingency was applied to action verbs.

The nature of the stimulus prompt

A prompt is defined as the stimulus actions or directions provided to a writer to elicit and focus the writing task (Meredith & Williams, 1984; Prater & Padia, 1983b). A literature review shows that there are various ways of stimulating creative writing. For example some prompts such as reading selections, trips, story outlines, newsworthy events, paintings, television, dance, cartoon characters, music and holidays (Graham, 1982) film (Berry, 1958; Parkins, 1963; Witty & Martin, 1957), gross stimuli such as toys, books and pictures have led to more successful writing than others. The latter include story titles (Carlson, 1963; Howell, 1956), objects and single words (Marshall, 1960), phrases (Campbell & Willis, 1978), brand names (Fagerlie, 1975), imaginary rather than actual experiences (Edmund,

1958a, 1958b, 1959; Wyatt, 1961), pictorial stimuli (Getzels & Jackson, 1960; Grandholm, 1983; Osterweis, 1968; Staudacher, 1968) choice, such that performance was improved when the topic was of personal interest to the student (Clark, 1954; Edmund, 1959; May & Tabachnick, 1966; Sofell, 1929), first-hand, multisensory rather than vicarious, e.g., pictorial experiences (Littwin, 1935) and a variety of sensory media (Nelson, 1965). Sharples (1968) compared four stimuli (picture, poem, sound, and object) and found that more organized representations (picture and poem) resulted in slightly but not significantly better themes. He concluded that the teaching climate and student background were more important variables than the nature of the stimulus. Finally Deno et al. (1980) found little difference in performance when type of stimulus (topic sentence; picture stimulus; story starter) was varied.

The pictorial mode has received the most research attention. May and Tabachnick (1966) assessed which of three stimuli (organized, unorganized, or a choice of either) would result in the most creativity in children's writing. The findings varied with the grade and sex of the subgroup. Thus the authors recommended offering the class a mixture of the three stimuli. In this way individual differences in motivation could be allowed for. Children have been found to write fewer words when pictures are outdated (Poteet, 1979). Golub and Frederick (1970) found that the variables of student sex and picture content affected the complexity, quantity and quality of written composition more so than black-and-white concrete, colour concrete or abstract pictures.

Overall, there is little conclusive research on whether a particular stimulus is more effective in eliciting creative writing. There are few comparative studies of responses to various stimuli. Some researchers have even found no significant relationship between children's writing quality and stimulation source (Berry, 1958; Berse, 1974; May & Tabachnick, 1966).

Four factors were considered when selecting the various essay topics for this study. Firstly that the topic was appropriate for the target group, in this case adolescent males; that the topic was broad enough to provide reasonable scope for response; thirdly, whether the topic lent itself to emotional responses from the subject; and finally and most importantly whether the writing purpose was adequately addressed by the topic (Meredith & Williams, 1984).

The following sections will analyze in more depth, the intervention Experiment 2 adopted and the justification for doing so. Alongside evaluation, intervention is one of the most controversial areas in the field of written composition. One of the liveliest issues is whether the emphasis should be on punishers or positive reinforcers.

The obsession with error

It has been claimed that teachers have advanced degrees in flaw-detecting (Kantor, 1979) otherwise known as a preoccupation with incorrectness. For example, Wingfield (1975) has suggested there are five distinct techniques for dealing with error, all of which are negative. Terminology tends to be pathological, focusing on what students could not rather than could do (Kantor, 1979). This is perhaps a reflection

of traditions of past education.

Reasons have been offered for such teacher behaviour. For example the teacher lacks standards if every aspect is not marked (Bennett, 1983) or parents, educators and administrators feel perturbed if teaching occurs without visible intervention. But more importantly what are the effects of seeing "red" all over one's personal work? Some have claimed it results in numerous negative and few positive effects (Bennett, 1983). Others have suggested it fails to achieve the aim of improving composition (Koch, 1982; O'Dea, 1965). In addition it has damaged the teacher-writer relationship which could otherwise have been a fruitful interaction. Koch (1982) has suggested that excessive, insensitive criticism results in built-in fear and avoidance which leads to poor writing skills and destroys imagination. Similarly, Vargas (1978) has stated that because the comments are so punishing, red pencil and failure become associated over a long period of time, such that positive comments cause the student to flinch.

These assumptions derive support from oral language developmental research, which has stressed the importance of a climate of trust. Factors such as amount of speech, motivation to talk, the desire to communicate and experiment are positively influenced by adult reinforcement, in particular attention to content and ideas, irrespective of form (Cazden, 1972). Likewise Skinner (1957) suggested complex verbal forms are achieved by successive reinforcement of a child's first attempts to talk, and that criticism interferes with this process. Vargas (1978) also stressed that if parents responded to children's verbal errors, as teachers

did to writing, conversation would quickly be extinguished.

The majority of articles have been about the effect of criticism on the student's self-concept. It has been suggested that an evaluation strategy needs to assist the beginning writer rather than discourage and alienate him; and that a healthy self-concept and confidence, stemming from successes are essential to developing as a writer. Authors have suggested that the red pencil tends to be destructive especially of the self-concept of learning disabled students, who by virtue of their disability, are more vulnerable to negative feedback (Blake, 1976; Personke, 1968; Simms, 1983).

Many articles in the literature propound these views, although few provide much evidence.

Is negative criticism effective?

Research is divided on the issue of whether praise or criticism has a more beneficial effect on the child. A review of the literature reveals the following viewpoints.

Some studies have suggested that criticism in the form of negative comments leads to a decrease in creative expression, writing ability and motivation to improve writing (Biberstine, 1967; Seidman, 1968; Spaulding, 1963). Burton and Arnold (1963) found no improvement with moderate or intensive marking of errors.

Only one study supported the claim that criticism was effective. Buxton (1958) applied the interventions of (A) a positive summary comment irrespective of quality of work, with no grade or comment in the margins, and (B) thorough marking, grades, comments, discussion and revision. Intervention B led to better results than Intervention A. However this finding was not surprising, since praise was superficial in Intervention A and Intervention B was more comprehensive. Other criticisms were that neither intervention groups nor the amount of positive reinforcement were controlled.

Overall, research (Brimner, 1982; Schroeder, 1973; Stevens, 1973) has found no difference between positive and negative written feedback. Where there has been a slight difference, studies have found that negative written comments affect mechanics positively, but attitudes negatively. They found that positive feedback improves attitudes, which is considered more important to written composition, in particular creativity, than mechanical improvements. These findings can be seen as being influenced by two factors. Firstly any comment can be interpreted as a sign of teacher interest. Thus both negative and positive comments function as attention and so could be reinforcing and thus effective. Secondly the studies tended to be methodologically weak. In particular, separate negative and positive feedback groups were not tightly controlled. A group receiving only negative feedback at times also received encouragement. This could mellow the effect of negative reinforcement and result in no significant differences between positive and negative reinforcement groups, which tended to be the case.

The majority of research has consistently supported the position that positive feedback is more effective than negative feedback in improving written composition. Perhaps the most well-known study was that of Taylor and Hoedt (1966) who assessed the effects of praise and criticism on the quality and quantity of creative writing in fourth graders. The praised group produced significantly more work, expressed more favourable attitudes, and were more highly motivated and independent.

In Nikloff's (1966) study teachers emphasized either low (ideas and originality) or high (spelling, neatness and conventions) standards. Students whose teachers emphasized low standards, produced slightly more words and ideas, fewer mechanical errors, and a higher overall quality of writing.

Clarke (1969) appears to have been the only researcher to assess the effect of number and type of written teacher comments on attitude. She concluded that the number of comments (range 2-5) had little effect and that purely negative comments resulted in lower scores than completely positive comments, in the areas of reinforcement, satisfaction and confidence. However these results were based on only one essay.

Gee (1971, 1972) assessed the effects of either praise, negative comment or no written comment on the compositions of high, middle and low ability eleventh graders. Positive comments resulted in significantly more favourable attitudes toward writing, increasing composition length and maintaining number of T-units.

The most consistent finding in the literature has therefore been that children learn to write better when praised than when criticized. Many studies however have shared similar methodological weaknesses, namely: the nature of the written comment, which is often ambiguous; the lack of interscorer reliability; poor design; small sampling of written products; the undefined nature of negative or positive feedback and lack of long term evaluation.

Positive reinforcement

This principle is based on the notion that positive reinforcement made contingent on a response tends to increase both the frequency of that behaviour and the probability that the response will occur again in the future. Vargas (1978) has taken a behavioural view of writing. She suggested that instead of the traditional method of teaching writing via correction with a red pen, so that writing became an aversive experience, the focus needed to be on contingencies manipulating variables that operated in the writing situation. Likewise, Emig (1977) suggested that writing itself uniquely offered both feedback and reinforcement, because the process could be immediately and visually seen in the written product.

Empirical evidence for the assumption that enthusiasm, acknowledgement of success however small, and encouragement constituted a better atmosphere for written composition, than anger and fear due to punishment has already been presented. In addition, Lysakowski and Walberg's review (1981) pointed to the effectiveness of rewards or positive reinforcement (both primary and secondary) on classroom learning in studies

spanning 20 years. The effects were consistent from kindergarten to university, across socioeconomic status, race, type of school and community.

The form of positive reinforcement

Positive reinforcement can be applied through a wide range of procedures: oral and/or teacher comments emphasizing social reinforcers (McKessar, 1977) in combination with corrective feedback (Barringer & Gholson, 1979; Bloom & Bourdon, 1980; Kulhavy, 1977; Solomon & Rosenberg, 1964; Walker & Buckley, 1972); public posting of performance data (Van Houten studies); token reinforcement programmes (Brigham et al., 1972; Maloney & Hopkins, 1973); group and individual contingencies; peer and self-reinforcement, and conferences.

The traditional form of evaluation has been the use of letter or numeral grades. Researchers have tended to criticize their efficacy (Dilworth & Reising, 1979; Dusel, 1955; Judy, 1973). Criticism has included that grades do not communicate to the student the areas which need improvement, nor encourage him to improve. Although there is widespread support for this view (Calabrese, 1982; Irmscher, 1979; Kehl, 1970) it is based on little empirical evidence.

An alternative procedure to giving grades is providing written commentary. The few studies which have compared written comments with lettergrade efficacy, have concluded that written comments are more effective than grades. Green (1968) found that symbols were less effective when the focus was on ideas rather than mechanics, in which case written comments should be used.

McKessar (1977) assessed the effect of numerical ratings and traditional written comments; team points; nonevaluative verbal comments; and a package of the latter three reinforcers, on the number of words and number of spelling and grammar errors, and subjectively rated quality, of Form I students' compositions. He found that package was the most successful and numerical ratings were the least effective.

Research has suggested that comments should be neither general, truncated, nor false (Calabrese, 1982; Koch, 1982; Simms, 1983). Rather they should be meaningful and creative (Kowalski, 1983) and positive and specific (Green, 1968; Kantor, 1979).

In Experiment 2 social reinforcement was presented through written and verbal content feedback. Verbal content feedback has been defined (Jerram, 1985) as occurring when the writer is informed about the effect (e.g., interest, enjoyment and excitement) the writing has had on the reader. In written content feedback this message from the reader is actually written back to the writer. Verbal content feedback, reinforcing and elaborating on the written content feedback, has been shown to occur within a responsive learning context (Glynn, 1980a, 1980b, 1980c).

The responsive learning context

Much commentary to date has emphasized that positive reinforcement and feedback are best delivered within the framework of a "conference" (Bissex, 1982; Carnicelli, 1980). However there has been little empirical research on its effectiveness.

The underlying principle behind the success of the conference is the concept of audience. By acknowledging the variable of audience researchers (King, 1978) have challenged the assumption that meaning resides in the written product alone. This view is presented pictorially in Figure 3.

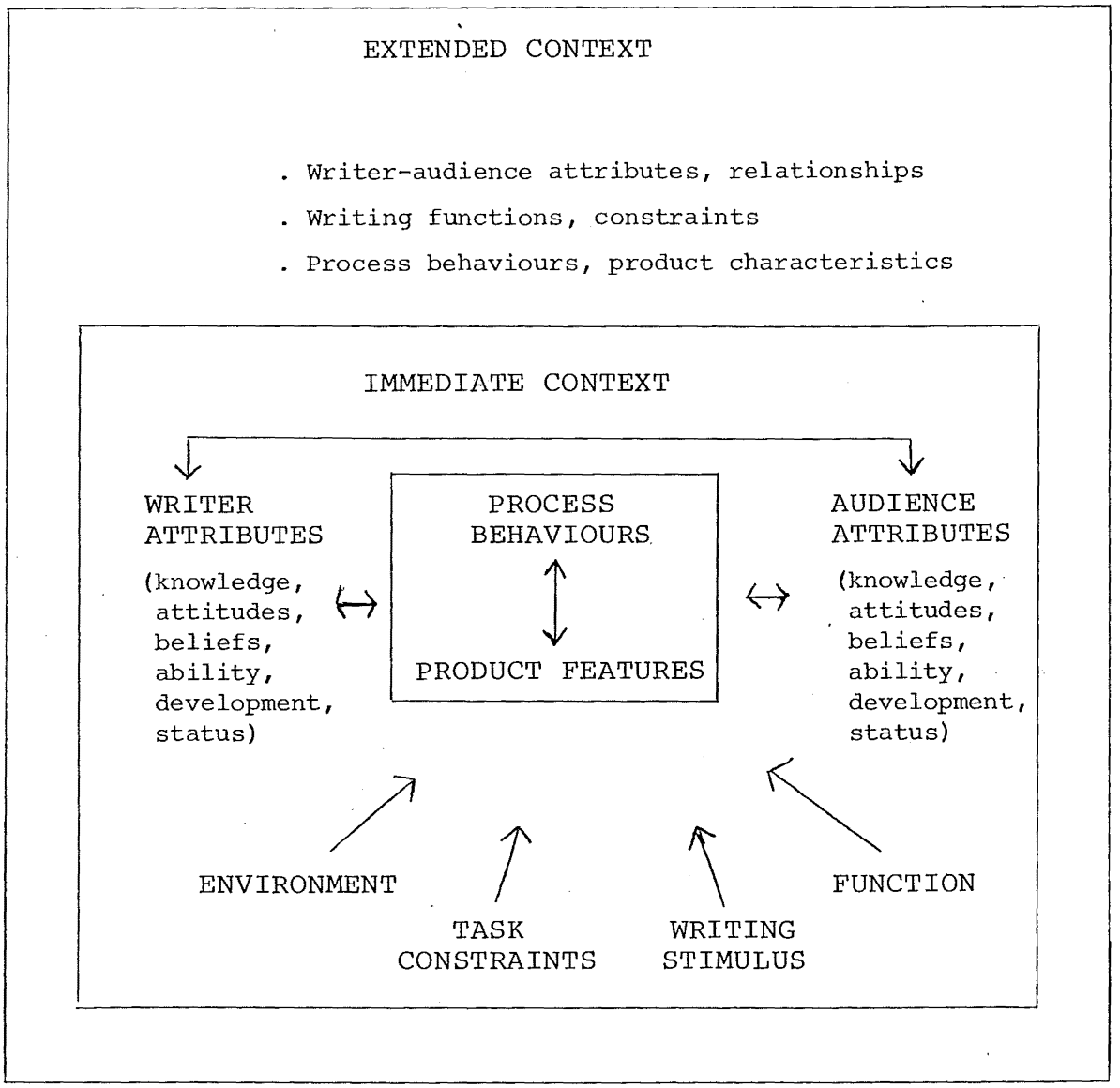
Insert Figure 3 here

Audience response to the writer's work has played a central role in a behavioural approach to writing. Skinner (1957) acknowledged the audience variable from both the side of the writer and reader. Researchers have suggested that writers need to develop a sense of audience (Beach & Bridwell, 1984; Golden, 1980; Kantor & Perron, 1977; Rosen, 1973; Sager, 1977). Hence they must be able to make adjustments and choices in writing which take account of the audience for whom the writing is intended. Readers too must provide an authentic audience and purpose for children's writing (Barenbaum, 1983; Beach & Bridwell, 1984; Golden, 1980; Jerram, 1985; Vargas, 1978).

Hagaman (1980) suggested that the major causes of ineffective teacher responses are failure to acknowledge what a writer conveyed, to involve the writer in advice given for revision, and the reasons for their own reactions. Of Bronfenbrenner's (1979) four properties of the ecological environment, his first proposition was relevant to the writer-reader relationship. This proposition concerned a "primary developmental context" which he defined as "one in which the child can observe and engage in ongoing patterns of progressively more complex activity jointly with or under the direct guidance of persons who possess knowledge and skill not yet

FIGURE CAPTION

FIGURE 3. Composition research variables. From New Directions in Composition Research (p.6) by R. Beach and L.S. Bridwell, 1984, New York: The Guilford Press.


EXTENDED CONTEXT
. Writer-audience attributes, relationships
. Writing functions, constraints
. Process behaviours, product characteristics
IMMEDIATE CONTEXT
WRITER ATTRIBUTES
(knowledge, attitudes, beliefs, ability, development, status)
PROCESS BEHAVIOURS
PRODUCT FEATURES
AUDIENCE ATTRIBUTES
(knowledge, attitudes, beliefs, ability, development, status)
ENVIRONMENT
TASK CONSTRAINTS
WRITING STIMULUS
FUNCTION

acquired by the child and with whom the child has developed a positive emotional relationship" (p.845). He provides persuasive and consistent evidence in support of this proposition.

More recently the work of Glynn (1985a, 1985b, 1985c) has expanded on the ideas of Bronfenbrenner (1979) and emphasized the importance of a "responsive social context" in the area or learning. Responsive learning contexts are characterized by four features (Glynn, 1985a). Firstly they should promote initiations by the learner. This can be achieved, for example, by providing various materials to engage the learner, and by delaying and reducing direct intervention by the teacher. Secondly they should provide an opportunity for the learner to engage in a shared activity with a skilled person with whom there is a positive social relationship. This relates to Bronfenbrenner's primary developmental context. Reciprocity or mutual influence is the third feature, requiring that writers and readers not only share the learning task but also modify each other's behaviour. Amount (regular) and type (responsive) feedback is the final requirement of a responsive learning context, a feature often inadequately addressed in the school classroom. Glynn (1985a, 1985b) supported the need for these four features with empirical evidence and demonstrated their influence and importance within a responsive learning context.

Glynn (1985b) has also commented on ways of providing a responsive social context for writing. Like Vargas (1978), he has stressed the importance of an audience response to the writer, indicating to the writer that his work has an impact on the reader. A responsive social context also means

that the audience response is not in the form of corrective feedback. Sharing and reciprocation, whether between teacher and child or parent and child, has to focus on the context or ideas and meaning the writer is trying to communicate, not on mechanical errors. Glynn (1985b) presented details of two studies in particular (Arndt, 1980; Jerram, 1985) which not only provide support for the effectiveness of applying a responsive social context to writing, but demonstrate how this was done.

In Experiment 2 positive written and verbal content feedback were provided within a one-to-one responsive social (non-evaluative) context, such that the audience responded in a non-corrective, personal, warm and positive way to the content and ideas of the writer's work. Firstly only the headmaster, less significant than the classroom teacher, provided social reinforcement and written content feedback (Intervention I). Although this was expected to lead to improvements over baseline, stronger effects were expected to occur in Intervention II. In this phase, the classroom teacher - in an even closer reciprocal relationship than the headmaster with the children - provided social reinforcement, written content feedback and in addition built on this through verbal content feedback. Intervention III was similar to Intervention II, except for the increased length of time in which verbal content feedback was reciprocated with each individual child.

The responsive social context was also relevant to Experiment 1 which implemented correspondence training (Christie & Ballard, 1983; Glynn, 1985a; Risley, 1977).

This intervention procedure required that the reader and writer confer prior to writing, to discuss the level of the target behaviour to be achieved. Even when the composition had been written, the student was again involved within a reciprocal learning context. The writer stated whether correspondence between the verbal statement and target behaviour had been achieved. The audience responded with praise and support, and also feedback as to the nature of the correspondence.

Feedback and social reinforcement made contingent on the content not form of written compositions was the focus of Experiment 2. A comparison was also made of the effects of action and theme picture prompts. The population was mildly retarded boys, an alternating treatments design was used and 24 collateral measures were assessed.

Experiment 2 examined four experimental questions:

(1) Did action picture prompts increase action verbs?

(2) Were stories containing higher numbers of action verbs rated as more creative?

(3) Did action picture prompts lead to more improvements on collateral behaviours than theme picture prompts?

(4) What were the effects of written/verbal content feedback and social reinforcement delivered within a one-to-one responsive learning context on the target behaviour, collateral behaviours and holistic scores?

METHOD

Subjects and Setting

The study took place in a special classroom in a

residential school for mentally retarded and emotionally disturbed boys. Data were obtained from four white male students (C.A. range 14 years 3 months to 16 years 0.5 month; IQ range 61-83) although all class members were exposed to treatment procedures. Further subject details are presented in Table 16. These subjects were mildly mentally retarded

Insert Table 16 here

according to the AAMD criteria (Grossman, 1983). Sex, intelligence, age, academic ability, race and socioeconomic status, and climate of school and community were controlled.

Permission to conduct the study was obtained from the principal of the school and the Department of Education. Persons involved in the study were the classroom teacher who volunteered to participate in the study, the headmaster, and experimenter who liaised with these people, discussed problems that arose, and xeroxed and analysed the stories.

Stimulus materials

The subjects were required to write stories in response to stimulus prompts in the form of 12 x 12 black and white pictures, photocopied from books at the public and school libraries. These were attached to the exercise book as required, and the essays were written below. The pictures were approved by the teacher and a photo-journalist as suitably depicting either action or theme scenes. Interestingly, it was difficult finding action pictures in books. They were scarce and often did not depict enough action. All subjects received the same picture for that session, and the two prompt types, action and theme, were randomly assigned (by

Table 16

Subject details in Experiment Two

Subjects	Age (years, months)	I.Q. (WAIS-R)	Reading Age (years)	Drugs
Bevan	16,0.5	63-73 (1982)	10-11	-
David	15,11.5	73-83 (1981)	8-8½ (Burt)	Nuelin SR 250 mg 1 tab twice daily. Intal spincap 1 capsule 3 x daily. Ventolin spandette 1 tap. at bedtime. Ventolin inhaler as necessary.
Aaron	14,10.5	70-80 (1981)	pre-reader (Burt)	-
Richard	14,3	61-71 (1981)	7-8 (Burt)	Veractil 25 mg 3 x daily

tossing a coin) to minimize carryover effects.

The decision to use picture prompts to stimulate story ideas was supported by the teacher's comment that subjects would be unmotivated otherwise. The topics of these pictures can be found in Appendix B.

Performance variable

The intervention of positive individual reinforcement and feedback, in the form of either written or both written and verbal comments, were made solely on creative responses in the written composition. A minimum of two but no more than three positive comments (if the stories were very good) were made per essay and a minimum of two sections (words or phrases) but no more than three (if ideas or use of language was very good) were circled per essay. The marker was provided with the "Carlson Analytical Originality Scale Key for scoring Original Stories" (Carlson, 1973) as a guideline on what constituted a creative response. The marker was also provided with a set of rules:

(a) No negative comments were to be written or verbalized.

(b) The focus was on the ideas and content written, with attention to creative responses. No comments were to be made about grammar, length or spelling.

(c) The comments were personalized and included the boy's name.

(d) Traditional comments, such as "good", "good work" were not to be used.

(e) Each subject's story was to be compared only to other stories written by that student, hence intrasubject marking. There was to be no intersubject marking,

nor marking based on absolute or normative data.

(f) No tangible rewards (such as tokens, free time or progress graph) were to be associated with the comments.

As in Experiment 1, in addition to measuring the performance variable, in this case creative responses as measured by the creativity score, measures of collateral behaviours were taken.

Experimental design

An alternating treatments design (Barlow & Hayes, 1979) with replication across four subjects was utilized. In this design, different treatments, whether therapists, settings, times or stimulus prompts, are systematically varied and counterbalanced across stimulus conditions within the same phase. The two conditions which were compared were (a) action oriented pictures and (b) theme pictures. In addition, the effect of positive reinforcement and feedback on creative responses in the written composition was assessed between baseline and intervention phases. The quality and quantity of feedback in intervention increased across phases. In Intervention I only written comments were made by the headmaster. In Intervention II the classroom teacher made the comments and supplemented written comments with verbal ones in a conference of one minute per subject. In Intervention III the same procedure as in Intervention II was used, but the conference was extended to two minutes (see Table 17).

Insert Table 17 here

Table 17

Experimental design for Experiment Two

APRIL		MAY - AUGUST	
Baseline	Intervention I	Intervention II	Intervention III
	Written comments and circling of good sections	Written comments and circling of good sections	Written comments and circling of good sections
No comments or circling of good sections	Headmaster wrote comments	Class teacher wrote comments	Class teacher wrote comments
		Individual one minute conference	Individual two minute converence
Session			
1 10 11	16 17	24 25	32

Procedure

Baseline Three samples of writing were obtained weekly on Mondays, Wednesdays and Fridays between 9 and 10.30 a.m. Having distributed the exercise books with that day's picture prompt attached, the teacher made the following general announcement: "Look at the picture in your book. You have half an hour to write a good story about it." The teacher was told to act as he normally did and keep his behaviour consistent between baseline and intervention. If the subject asked the teacher how to spell a word, it was written at the top of the essay. Apart from this contact, the teacher was instructed to make no other comments to individuals or the class. During baseline no comments were written at the end of the story. Prior to the study, the teacher never gave written comments, but he did write the correct spelling over an incorrect word. He was given the option of stopping this practice during the study, or if continuing it, to do so consistently across baseline and intervention. He chose the former option. At the end of each 30 minute session, the teacher asked the boys to stop writing and collected the books.

Intervention I The same procedure as in baseline occurred. In addition when the books were collected they were sent to the headmaster (on Tuesday, Thursday and Friday afternoons) for a written comment and circling of good sections. In this way the boys received relatively immediate feedback and were able to read the comments before they started their next story on Mondays, Wednesdays and Fridays. The comment, written in blue pen below the essay, always included the boy's name and was signed by the marker's

initials. The experimenter did not write the comments since she was not a significant other to the boys and her comments may have had less effect than the teacher's.

Intervention II The same procedure as in Intervention I was utilized, except this time the classroom teacher did the commenting and circling. In addition to written comments, before writing their next story, the teacher went to each boy's desk and in a one minute conference discussed the written comments and circled sections with each boy and praised him verbally and with smiles. The conference again only focussed on creative responses.

Intervention III This phase utilized similar procedures as in Intervention II, except in this phase the conference was extended to two minutes. Because this involved more of the teacher's time, he assigned Bevan (one of the subjects) the role of giving those who asked, the correct spelling of words. The teacher verified that Bevan could carry this out satisfactorily.

Because of time constraints, a maintenance phase was unfortunately not implemented.

Dependent variables

In Experiment 2 each essay was analyzed by measures of both the target behaviour (creative responses) and collateral behaviours. These measures were assessed in exactly the same way as in Experiment 1, except for the measures of capitalization and action verbs and the addition of the contact component. The percentage correct for capitalization was calculated by dividing the total number of letters correctly capitalized by the number of letters which should have been capitalized. Unlike Experiment 1, it

was not divided by the number of letters inappropriately capitalized as this error occurred so frequently that the score would be distorted. The percentage total action verbs was calculated by dividing the total action verbs by the total number of words written and multiplying this by 100. It was not divided by total verbs, as occurred in Experiment 1, as total verbs were so few as to make it unfeasible.

The contact component applied to Experiment 2 only. It concerned contacts made with the teacher during the half hour in which the essays were written. The percentage of words asked how to spell measure was derived by dividing the number of words asked how to spell by the total number of words written and multiplying the quotient by 100.

Holistic scoring

Undergraduate scorers marking Experiment 2 essays received a packet of 16 essays. The last six essays of each boy's baseline and Intervention II phases were selected and photocopied three times. Thus the same essay was marked by three different scorers. The subject's mean holistic score per essay was obtained by averaging the independent ratings of the three scorers. The subject's mean score for the six essays selected for evaluation from each phase was averaged to give a mean holistic score per baseline and intervention phase per subject. Other than these differences, holistic scoring was carried out exactly as in Experiment 1.

Reliability

Reliability procedures were identical to Experiment 1.

(a) Analytic scoring. Table 18 shows that interscorer

Insert Table 18 here

agreement was overall very high, the highest mean being 100% (total words written and words asked how to spell) and the lowest 81.97% (creativity score). Except for the latter result, all measures assessed, received reliabilities in the 90s. The range varied from no difference (100%-100%) to a 50% difference (50%-100%).

(b) Holistic scoring. Of the 48 essays selected for holistic scoring, 75% (36 essays) were rated reliably, scores differing no more than one or two points. Of the 25% (12 essays) that were unreliably rated, 15% (7) received scores ranging by three marks and 6% (3) received scores ranging by four different marks. On the remaining two essays the scores ranged as wide as 3-8 and 1-7 on the 0-10 scale.

(c) Written content feedback. Regarding quantity, the written comments and circled sections were counted to assess if readers provided equal amounts of feedback across subjects and phases. In all cases, they were well controlled and reliably delivered in line with instructions given. Markers also reliably controlled the quality of comments delivered to subjects across phases. Written content feedback consistently included elements considered important in teacher-pupil conferences (Jerram, 1985; Searle & Dillon, 1980). Story content was consistently focussed on to the absolute exclusion of form.

Table 18

Summary of interscorer reliability for Experiment 2

Measure	Mean	Range
Handwriting score	93.47	83-100
Total words written	100	100-100
Total sentences	98.90	67-100
Total T-units	96.13	67-100
Vocabulary diversity	99.83	95-100
% correct capitalization	95.9 97.23	50-100 50-100
% correct punctuation	99.03 92.90	80-100 50-100
% spelling inaccuracy	96.60	60-100
Vocabulary score	98.60	75-100
Word usage score	96.20	90-100
% mature words	97.37	67-100
Grammar errors	91.90	50-100
% total nouns	98.90	75-100
% total action verbs	94.70	50-100
% total adjectives	93.37	50-100
% total adverbs	94.27	50-100
% diff. nouns	98.87	75-100
% diff. action verbs	94.70	50-100
% diff. adjectives	94.47	50-100
% diff. adverbs	94.27	50-100
Mean T-unit length	96.17	63-100
Thematic maturity	91.2	67-100
Creativity score	81.97	50-100
Words asked how to spell	100	100-100

Statistical procedures

A repeated measures ANOVA was used to analyse Experiment 2. The action, theme, and total scores were entered separately into the one way repeated measures design. (In this case, the four levels of the repeated measures factor corresponded to the baseline, and three treatment phases). The same series of planned contrasts were employed as in Experiment 1.

The differences between thematic and action condition score levels, and patterns of scores on the dependent measures were also examined. A two factor ANOVA with repeated measures factors of phase and type of picture was used. Differences in the TYPE main effect indicates different rates of responding to the pictures. Differences in the TYPE x PHASE intervention term indicates a different effect of treatment for the two picture conditions.

As with Experiment 1, repeated measure t-tests were used to assess differences in holistic ratings. Comparisons between picture conditions at baseline, and at follow-up were also tested for in this way.

RESULTS

Target Behaviour

The intervention was applied to the target behaviour of creative responses. As shown in Table 19 and Figure 4,

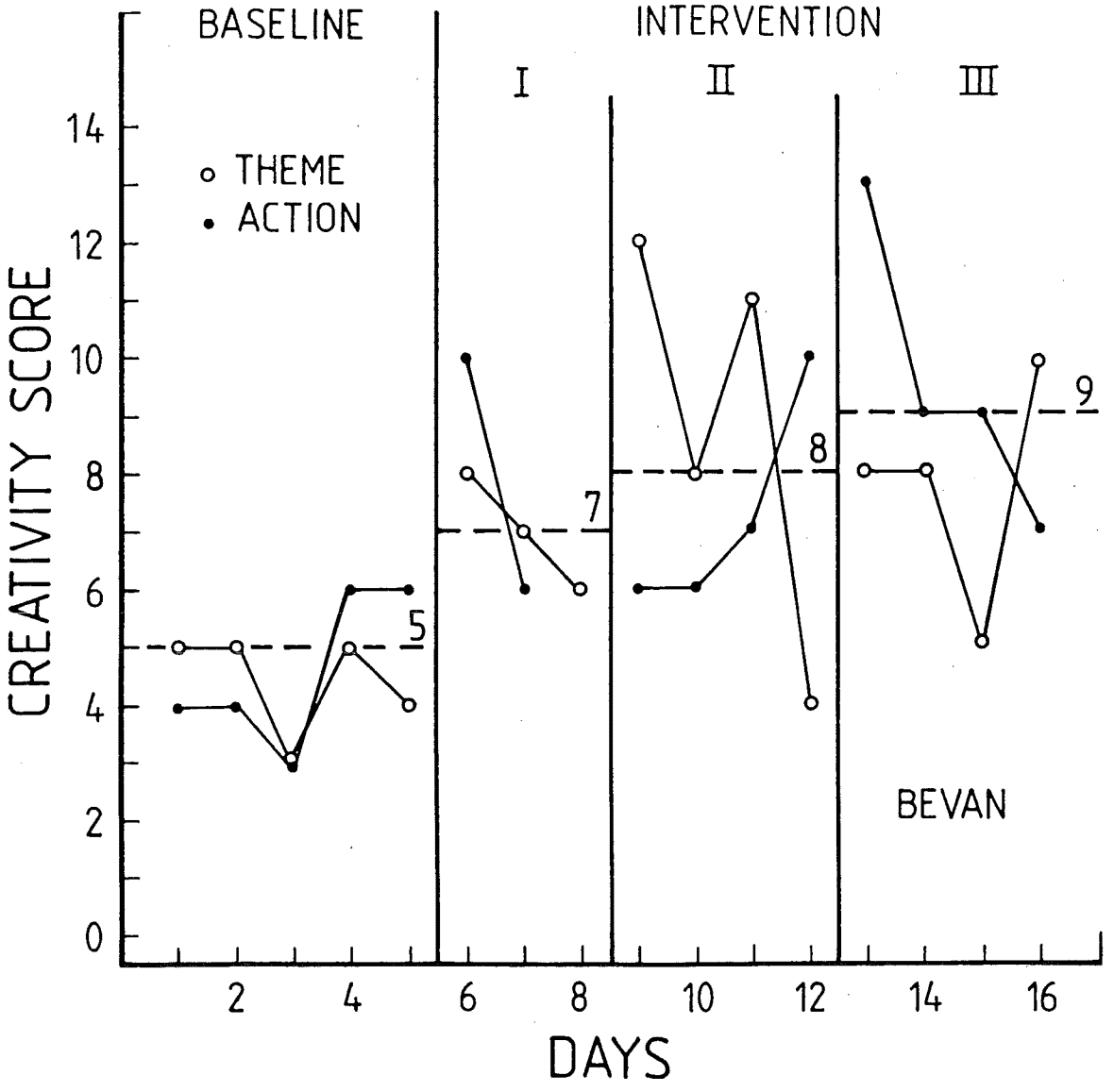
Insert Table 19 and Figure 4 here

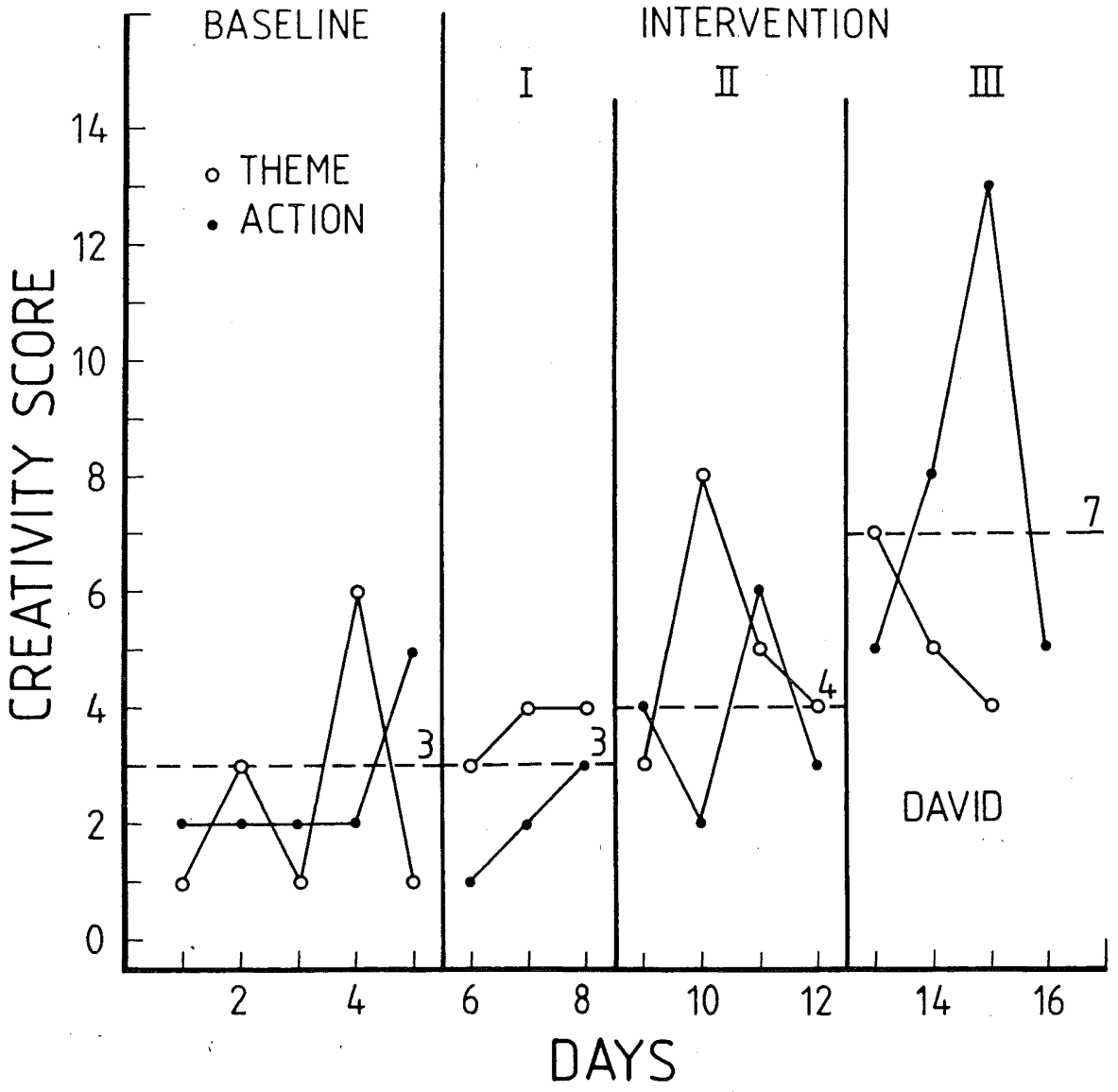
Mean scores across subjects and experimental conditions

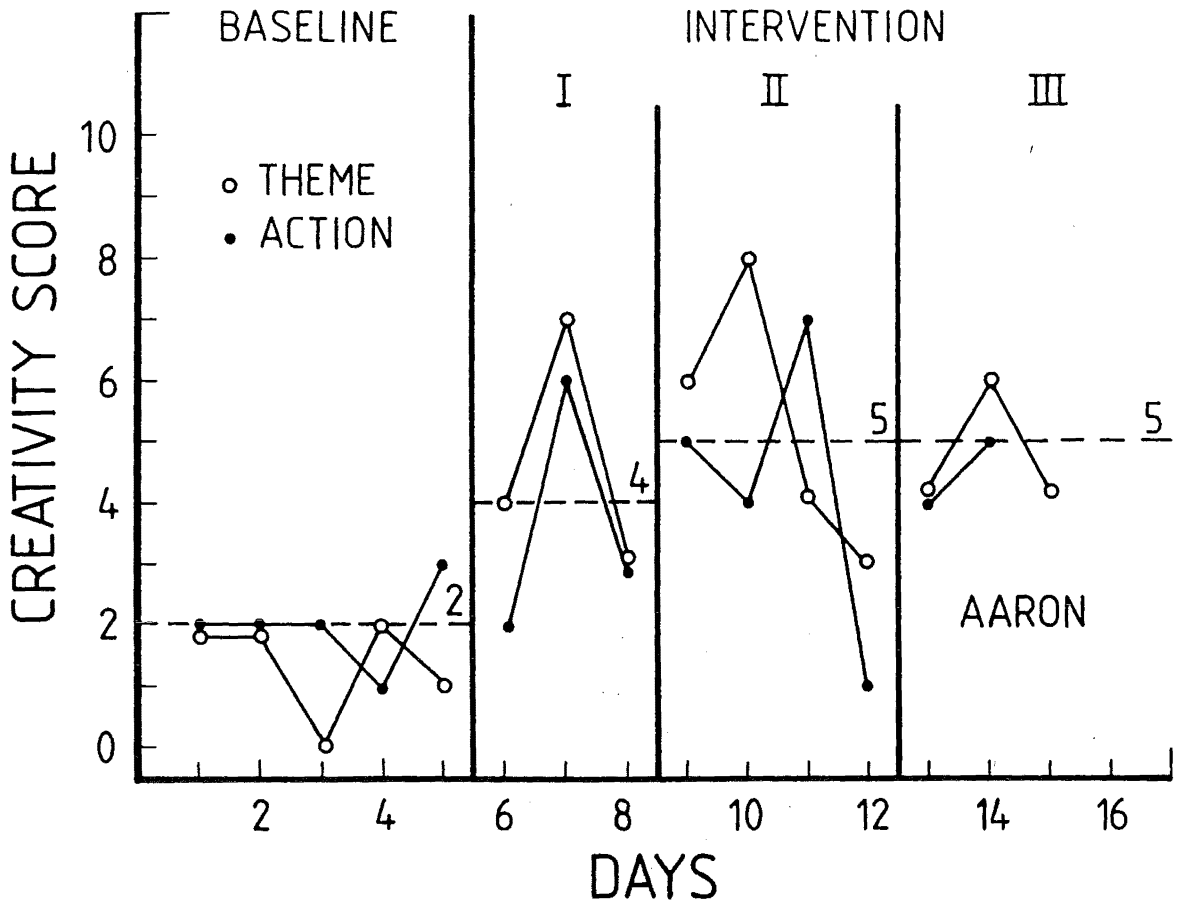
Subjects	Baseline			Intervention I			Intervention II			Intervention III		
	Action	Theme	Mean	Action	Theme	Mean	Action	Theme	Mean	Action	Theme	Mean
1. Creativity												
Bevan	5	4	5	8	7	7	7	9	8	10	8	9
David	3	2	3	2	4	3	4	5	4	8	5	7
Aaron	2	1	2	4	5	4	4	5	5	5	5	5
Richard	3	3	3	4	3	3	5	5	5	5	7	6
2. Thematic												
Bevan	2	2	2	3	3	3	4	4	4	3	3	3
David	2	2	2	1	3	2	2	3	3	4	3	3
Aaron	1	2	2	3	2	3	3	4	3	3	2	3
Richard	2	2	2	2	2	2	2	2	2	2	3	3
3. Handwriting												
Bevan	5	4	5	6	6	6	6	5	5	5	5	5
David	6	7	6	5	6	5	6	5	5	7	6	6
Aaron		6	6				6	7	6	7	7	7
4. Words												
Bevan	37	32	35	68	61	64	41	41	41	40	35	37
David	17	14	16	19	24	22	21	25	23	40	40	40
Aaron	15	13	14	25	23	24	21	21	21	24	19	21
Richard	16	15	16	15	10	13	14	17	16	23	23	23
5. Sentences												
Bevan	4	4	4	5	4	4	4	4	4	4	3	3
David	3	3	3	2	4	3	3	3	3	5	3	4
Aaron	1	2	2	3	4	3	3	3	3	4	2	3
Richard	2	2	2	2	1	2	2	3	2	3	3	3
6. T-units												
Bevan	4	5	4	7	5	6	5	4	4	4	4	4
David	3	3	3	2	4	3	3	3	3	6	3	5
Aaron	1	2	2	3	4	3	3	3	3	4	2	3
Richard	2	2	2	2	2	2	2	3	2	3	3	3
7. New words												
Bevan	39	41	40	27	37	33	29	29	29	29	25	27
David	47	53	50	26	26	26	29	37	33	27	26	27
Aaron	45	34	40	36	29	33	26	33	29	19	22	21
Richard	54	54	54	39	28	34	28	38	33	20	23	21

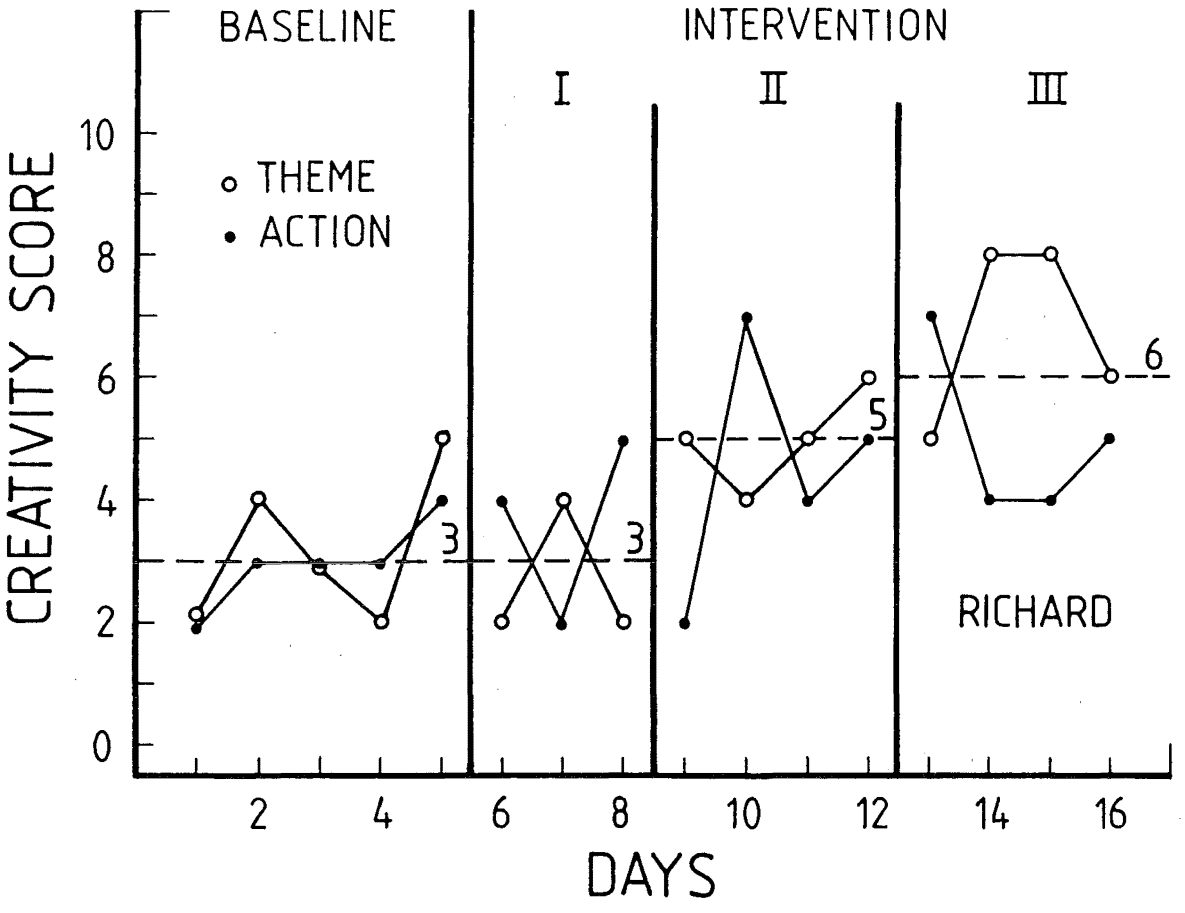
FIGURE CAPTION

FIGURE 4. Mean creativity scores for Bevan, David, Aaron, and Richard across baseline and intervention phases.


BASELINE
INTERVENTION
I
II
III
THEME
ACTION
CREATIVITY SCORE
DAYS
BEVAN
5
7
8
9
0
2
4
6
8
10
12
14
16


BASELINE
INTERVENTION
I
II
III
THEME
ACTION
CREATIVITY SCORE
0
2
4
6
8
10
12
14
3
3
4
7
DAVID
2
4
6
8
10
12
14
16
DAYS


BASELINE
INTERVENTION
I
II
III
THEME
ACTION
CREATIVITY SCORE
10
8
6
4
2
0
2
4
6
8
10
12
14
16
DAYS
2
4
5
5
AARON


BASELINE
INTERVENTION
I
II
III
THEME
ACTION
CREATIVITY SCORE
10
8
6
4
2
0
3
3
5
6
RICHARD
2
4
6
8
10
12
14
16
DAYS

the mean creativity score increased from baseline to Intervention I for two boys (Bevan and Aaron) and increased for all subjects in Interventions II and III. The increase from Intervention I to II was statistically significant [$F(1,3) = 25.00$, $p < .015$] as was the increase from baseline to Intervention III [$F(1,3) = 147.00$, $p < .001$].

The comparison between creativity scores on action and theme picture prompts showed little difference. Action pictures resulted in slightly higher mean creativity scores than theme pictures in baseline and Intervention III, but there was little difference between the two picture prompts in Intervention I. In Intervention II, theme pictures produced a slightly higher mean creativity score than action pictures.

Statistical analysis for action picture prompts showed an overall significant effect for the creativity score [$F(9,3) = 7.26$, $p < .009$] and the improvement from baseline to Intervention III was statistically significant [$F(1,3) = 25.00$, $p < .015$]. Similarly there was an overall significant effect for theme picture prompts on the creativity score [$F(9,3) = 13.65$, $p < .001$] and the increase from baseline to Intervention III was also significant [$F(1,3) = 225.00$, $p < .001$].

For the mean thematic maturity scale (score 1-10) as shown in Table 19 all subjects' scores increased from baseline (mean score of 2) across the interventions (mean score of 3). For Richard the increase did not occur until Intervention III, for David Intervention II, and for Bevan and Aaron still earlier at Intervention I. There was little

difference whether the picture prompt was action or theme in nature, for this measure.

Collateral behaviours

Mechanical component. The mean handwriting score as shown in Table 19 was applicable for three boys (Bevan, David and Aaron). The results showed that handwriting remained relatively constant in quality throughout baseline and intervention phases. There was also little difference whether the prompt was action or theme.

Productive component. The mean number of words (Table 19) increased from baseline to Intervention I. It decreased slightly in Intervention II (for Bevan and Aaron) though still remaining above baseline levels, and this level remained in Intervention III. The other subjects (David and Richard) increased their output in Intervention III. The greatest increase was in David's mean score from 16 in baseline to 40 in Intervention III, a 150% increase. Overall, the action pictures resulted in slightly more words written than theme picture prompts.

The mean number of sentences (Table 19) increased from baseline to Intervention III for three subjects (David, Aaron and Richard). Bevan remained at the same level from baseline to Interventions I and II and decreased slightly in Intervention III. Type of picture prompt had little effect on the mean number of sentences.

The mean number of T-units (Table 19) increased slightly for all subjects from baseline to Intervention III, with the exception of Bevan. He increased during Intervention I but returned to baseline levels during Interventions II and

III. There was little difference in the mean number of T-units for action and theme prompts.

The mean percentage of new words (Table 19) decreased across phases, approximately 50% for David, Aaron and Richard and 30% for Bevan from baseline to Intervention III. This measure was overall statistically significant [$F(9,3) = 15.87$, $p < .001$]. The decreases between baseline and Intervention I [$F(1,3) = 10.83$, $p < .046$], between Interventions II and III [$F(1,3) = 11.31$, $p < .044$] and between baseline and Intervention III [$F(1,3) = 27.40$, $p < .014$] were all statistically significant.

The mean percentage of new words for action [$F(9,3) = 12.35$, $p < .002$] and theme [$F(9,3) = 9.55$, $p < .004$] picture prompts were statistically significant. The decrease in mean percentage of new words for action prompts between baseline and Intervention I [$F(1,3) = 30.94$, $p < .011$ and between baseline and Intervention III [$F(1,3) = 19.79$, $p < .021$] were statistically significant. The decreases between Intervention II and III [$F(1,3) = 20.09$, $p < .021$] and between baseline and Intervention III [$F(1,3) = 23.02$, $p < .017$ for theme picture prompts were also statistically significant. Overall there was little difference between type of picture prompt on this measure.

The mean vocabulary diversity score (Table 20) increased

Insert Table 20 here

from baseline to Intervention III for all boys, with a moderate increase for Bevan and David and a slight increase for Aaron and Richard. There was also a significant [$F(1,3) = 15.36$, $p < .030$] increase in this measure between baseline

Mean scores across subjects and experimental conditions

Subjects	Baseline			Intervention I			Intervention II			Intervention III		
	Action	Theme	Mean	Action	Theme	Mean	Action	Theme	Mean	Action	Theme	Mean
8. Vocabulary diversity												
Bevan	2.8	2.7	2.8	3.1	3.7	3.5	3.1	3.3	3.2	3.4	3.3	3.4
David	2.3	2.1	2.2	2.1	2.2	2.1	2.3	2.4	2.4	3.0	3.2	3.1
Aaron	2.4	1.9	2.2	2.6	2.4	2.5	2.6	2.6	2.6	2.5	2.5	2.5
Richard	2.3	2.4	2.4	2.2	2.2	2.2	2.3	2.5	2.4	2.4	2.7	2.5
9. Capitalization												
Bevan	97	93	95	100	100	100	82	100	91	92	100	96
David	100	100	100	100	93	97	88	92	90	89	83	87
Aaron	100	100	100	83	80	82	83	84	83	80	89	85
Richard	100	67	83	94	100	97	76	65	71	41	45	43
10. Punctuation												
Bevan	78	85	82	77	73	74	79	87	83	87	67	77
David	82	84	83	72	86	79	88	82	85	88	85	86
Aaron	100	75	88	57	61	59	74	81	78	80	83	82
Richard	50	20	35	42	33	38	0	15	7	8	0	4
11. Spelling												
Bevan	3	1	2	1	3	2	3	1	2	6	3	4
David	10	6	8	11	9	10	11	5	8	5	8	7
Aaron	18	23	20	15	14	14	18	12	15	10	14	12
Richard	8	10	9	6	10	8	5	13	9	18	14	16
12. Vocabulary												
Bevan	4	5	5	10	11	11	7	5	6	8	5	7
David	2	1	2	4	3	4	5	1	3	5	5	5
Aaron	2	1	1	4	2	3	1	2	1	2	1	1
Richard	1	1	1	2	1	2	1	2	2	3	3	3
13. Word usage												
Bevan	1	1	1	2	2	2	2	2	2	2	2	2
David	1	0	0	1	1	1	1	1	1	2	2	2
Aaron	0	0	0	1	0	1	1	1	1	1	1	1
Richard	1	1	1	1	0	1	1	1	1	1	1	1
14. Mature words												
Bevan	40	41	41	37	38	37	38	39	38	43	45	44
David	37	37	37	43	34	38	36	37	36	37	37	37
Aaron	20	31	26	36	24	30	27	29	28	28	25	26
Richard	30	36	33	36	25	31	29	36	32	26	33	29

and Intervention III for theme picture prompts. Overall there was little difference in this measure for theme and action picture prompts.

Conventional component. The mean percentage capitalization correct (Table 20) decreased across phases for all boys except Bevan. The decrease was moderate for David and Aaron (15%) however Richard's score decreased by 50%. Type of picture prompt had little effect on this measure.

Bevan and Aaron's mean percentage punctuation correct (Table 20) decreased slightly throughout baseline and intervention phases. Richard's score was low to begin with in baseline and Intervention I and decreased greatly in Intervention II and III. David remained at the same level across phases. There was no consistent trend in the effect of type of picture prompt on mean percentage punctuation correct.

The mean percentage spelling errors (Table 20) for David remained constant over baseline and intervention, as it did for Bevan and Richard except in Intervention III where their level increased. There was an almost 50% decrease in Aaron's percentage spelling errors from Baseline to Intervention III.

There was no relationship between picture type and percentage of spelling errors across phases. The increase in the mean percentage of spelling errors for theme picture prompts across Interventions II and III was statistically significant [$F(1,3) = 24.00$, $p < .016$].

Linguistic component. The first of the three semantic measures, the mean vocabulary score (Table 20) increased from baseline to Intervention I for all boys, decreased in Intervention II and then increased slightly in Intervention III to

higher than baseline levels for all but one boy (Aaron). Action picture prompts led to a slightly higher mean vocabulary score than theme picture prompts.

The mean word usage score (0 to 3 scale) increased moderately from baseline to Intervention III for David and Bevan, slightly for Aaron, with no change for Richard. (Table 20). The phase effect was significant [$F(9,3) = 4.40$, $p < .036$]. There was no difference in this measure whether action or theme picture prompts were used.

The mean percentage of mature words (Table 20) remained relatively constant over all phases for all boys. Bevan increased slightly in Intervention III. Overall the type of picture prompt had no consistent effect on mean percentage of mature words.

The first syntax measure, mean grammar errors per 100 words (Table 21) increased from baseline to intervention

Insert Table 21 here

phases across all four boys. This phase effect was statistically significant [$F(9,3) = 5.49$, $p < .020$].

There was no pattern according to type of picture prompt. For Bevan, both types showed equivalent error rates, however for the other three boys, results across phase and type were variable and showed no trends.

The mean percentage total nouns (Table 21) remained constant across phases for Bevan and Richard, it increased slightly for David and decreased slightly for Aaron. Overall, type of picture prompt did not differentially effect this measure for the subjects.

Subjects	Baseline			Intervention I			Intervention II			Intervention III		
	Action	Theme	Mean	Action	Theme	Mean	Action	Theme	Mean	Action	Theme	Mean
15. Grammar												
Bevan	3	3	3	2	4	3	5	4	5	4	4	4
David	4	11	8	5	13	9	7	18	12	15	15	15
Aaron	4	0	2	3	2	2	19	7	13	5	23	15
Richard	1	0	1	3	7	5	6	1	4	12	6	9
16. Total nouns												
Bevan	32	28	30	28	28	28	28	25	27	32	29	31
David	29	26	28	33	31	32	31	33	32	28	33	30
Aaron	26	30	28	26	24	25	28	23	26	24	26	25
Richard	27	23	25	33	23	28	30	31	31	20	29	25
17. Total verbs												
Bevan	11	5	8	10	7	8	7	8	7	8	11	10
David	10	3	6	9	8	9	9	7	8	13	7	10
Aaron	9	1	5	16	2	9	9	8	8	10	14	12
Richard	13	2	8	5	6	5	4	8	6	12	5	9
18. Total adjectives												
Bevan	9	10	9	13	8	10	8	13	11	4	9	7
David	5	11	8	9	0	5	7	5	6	8	3	6
Aaron	10	13	12	10	6	8	9	8	8	8	13	11
Richard	11	20	15	10	3	7	8	8	8	5	7	6
19. Total adverbs												
Bevan	3	4	4	1	5	3	5	4	4	1	5	3
David	4	8	6	5	7	6	2	3	3	3	5	4
Aaron	3	0	2	2	5	3	0	2	1	7	0	3
Richard	1	6	4	5	10	7	3	1	2	7	9	8
20. Different nouns												
Bevan	24	19	22	17	24	21	22	24	23	27	25	26
David	26	21	24	20	19	20	25	25	25	20	30	24
Aaron	26	25	25	22	17	20	24	23	24	18	24	22
Richard	26	23	25	33	23	28	27	31	29	19	29	24
21. Different verbs												
Bevan	8	5	7	8	6	7	7	8	7	7	10	9
David	10	3	6	9	8	9	6	7	7	13	6	10
Aaron	8	1	5	11	2	7	9	6	7	8	13	11
Richard	13	2	8	5	6	5	4	8	6	11	4	8

The comparison of baseline and Intervention III phases for mean percentage total action words (Table 21) shows increases in scores for all boys. This phase effect was significant [$F(9,3) = 4.50$, $p < .034$] as was the increase between Interventions II and III [$F(1,3) = 54.00$, $p < .005$].

The measure had overall significance for theme [$F(9,3) = 4.29$, $p < .039$] picture prompts and the increase between baseline and Intervention I was significant [$F(1,3) = 10.80$, $p < .046$]. Overall, the mean percentage for action picture prompts was higher than for theme picture prompts though less so in Interventions II and III. This type effect was significant [$F(3,1) = 19.37$, $p < .022$].

There was a slight decrease in mean percentage total adjectives from baseline to intervention phases for Bevan, David and Aaron, with a greater decrease for Richard.

In baseline, theme picture prompts led to more adjectives than action prompts, but this reversed in Intervention I. In Intervention II there was little difference between picture types on this measure. However in Intervention III theme pictures led to a higher mean level of adjectives for all boys except David. The measure had overall significance [$F(9,3) = 4.97$, $p < .026$] for theme picture prompts and the decrease in the percentage levels for theme pictures between Interventions I and II, was statistically significant [$F(1,3) = 32.11$, $p < .011$]. The phase by type effect was also significant [$F(9,3) = 7.23$, $p < .009$] and suggested that the patterns of change for the two types of prompts differed over the phases.

Bevan, David and Aaron maintained their level of mean percentage total adverbs (Table 21) between baseline and

Intervention III. Richard wrote more adverbs in Interventions I and III, but fewer in Intervention II than in baseline. With some exceptions, the mean total adverbs were higher for theme than action picture prompts across all boys and phases.

The level of mean percentage different nouns (Table 21) was more or less consistent across baseline and intervention phases. The only decrease tended to occur from baseline to Intervention I for two boys (David and Aaron). The increase in the level from Intervention I to II was significant [$F(1,3) = 10.80$, $p < .046$].

The effect of picture type on this measure varied. In baseline, action pictures led to more different nouns than did theme pictures, but this was reversed in Intervention III. There was little difference between picture prompts in Interventions I and II. Theme picture prompts had overall significance on this measure [$F(9,3) = 4.18$, $p < .041$].

All subjects bar Richard increased their mean percentage different action verbs from baseline to intervention and this phase effect was significant [$F(9,3) = 3.86$, $p < .050$]. The increase between Interventions II and III was also significant [$F(1,3) = 33.00$, $p < .010$].

Although the mean percentage level was highest for action picture prompts in both baseline and Intervention I, the decrease in Intervention II was significant [$F(1,3) = 13.36$, $p < .035$]. In Intervention III the mean percentage level was higher for action picture prompts for two boys (David and Richard) but lower than theme picture prompts for the other two boys (Bevan and Aaron). There was a significant type effect [$F(3,1) = 10.93$, $p < .046$].

Mean percentage different adjectives (Table 22)

Insert Table 22 here

decreased for all boys between baseline and intervention phases. This phase effect was significant [$F(9,3) = 5.77$, $p < .018$]. The decrease between baseline and Intervention III was also significant [$F(1,3) = 10.45$, $p < .048$].

Although theme picture prompts led to higher mean percentages in baseline, this trend reversed in Intervention I. In Interventions II and III the mean percentage ratios for action and theme prompts were similar for Aaron and Richard. For Bevan the theme prompt had a higher mean percentage level, contrasting with David where the reverse occurred.

These differences between phases for the two types resulted in a significant phase by type effect [$F(9,3) = 9.32$, $p < .004$]. There was overall significance in the measure for action picture prompts [$F(9,3) = 10.43$, $p < .003$] and the decreases in mean percentage levels between Interventions I and II [$F(1,3) = 24.00$, $p < .016$] and between Interventions II and III [$F(1,3) = 22.09$, $p < .018$] were significant. The measure also had overall significance for theme picture prompts [$F(9,3) = 6.90$, $p < .010$]. The increase in mean percentage levels between Interventions I and II [$F(1,3) = 32.00$, $p < .011$] was also significant.

Bevan and Aaron maintained a relatively constant mean percentage level of different adverbs (Table 22) across baseline and intervention phases. David's level however decreased by 50% between Interventions I and II, and he maintained this same level of responding in Intervention III. Richard

Table 22

Mean scores across subjects and experimental conditions

Subjects	Baseline			Intervention I			Intervention II			Intervention III		
	Action	Theme	Mean	Action	Theme	Mean	Action	Theme	Mean	Action	Theme	Mean
22. Different adjectives												
Bevan	8	10	9	10	7	8	7	11	9	4	7	5
David	5	10	7	9	0	5	7	5	6	6	3	4
Aaron	9	11	10	10	6	8	8	8	8	6	9	8
Richard	9	19	14	9	3	6	8	8	8	5	7	6
23. Different adverbs												
Bevan	3	4	3	1	4	3	4	4	4	1	5	3
David	4	7	6	5	7	6	2	3	3	3	4	3
Aaron	3	0	2	2	5	3	0	2	1	7	0	3
Richard	1	5	3	5	10	7	3	1	2	5	8	6
24. T-unit length												
Bevan	11	7	9	11	12	12	10	13	11	12	10	11
David	7	6	6	9	7	8	7	9	8	8	12	10
Aaron	11	7	9	8	6	7	7	8	7	7	9	8
Richard	7	9	8	9	7	8	8	8	8	7	8	8
25. Words asked												
Bevan	0	1	1	0	1	1	1	2	1	0	2	1
David	0	0	0	0	0	0	0	0	0	3	0	2
Aaron	0	0	0	2	1	2	0	2	1	0	0	0
Richard	7	9	8	9	6	7	8	10	9	7	12	10
26. Holistic score												
Bevan	7	7	7				8	9	8			
David	6	4	5				5	5	5			
Aaron	4	4	4				5	5	5			
Richard	6	5	5				5	5	5			

increased mean percentage levels from baseline to Intervention I, decreased in Intervention II, and then increased again in Intervention III.

Overall the mean percentage levels of different adverbs was higher for theme picture prompts than action picture prompts across all boys and phases.

The figures in Table 23 refer to the percentage

Insert Table 23 here

decrease when different parts of speech were subtracted from total parts of speech. This provided a measure of repetition of parts of speech. The mean decrease for nouns was 4% (range 0-12); for action verbs was 1% (range 0-2); for adjectives was 1% (range 0-2); and for adverbs was 0% (range 0-1).

The mean T-unit length (Table 22) increased from baseline to intervention phases for two boys (Bevan and David), while the other two boys (Aaron and Richard) maintained the same level of responding across phases.

Although the mean T-unit length was slightly longer for action than theme picture prompts in baseline and Intervention I, this reversed in Interventions II and III. The increase in mean T-unit length for theme pictures between Interventions I and II was significant [$F(1,3) = 27.00$, $p < .014$] and the decrease for action picture prompts between Interventions I and II was also significant [$F(1,3) = 25.00$, $p < .015$].

Contact component. Mean percentage words asked (Table 22) changed little across phases. Three of the boys asked very few words and maintained a consistent level of contact

Table 23

The decrease in percentage from total to different parts of speech

Subjects	Phases			
	Baseline	Intervention I	Intervention II	Intervention III
1. Nouns				
Bevan	8	7	4	5
David	4	12	7	6
Aaron	3	5	2	3
Richard	0	0	2	1
2. Action verbs				
Bevan	1	1	0	1
David	0	0	1	0
Aaron	0	2	1	1
Richard	0	0	0	1
3. Adjectives				
Bevan	0	2	2	2
David	1	0	0	2
Aaron	2	0	0	3
Richard	1	1	0	0
4. Adverbs				
Bevan	1	0	0	0
David	0	0	0	1
Aaron	0	0	0	0
Richard	1	0	0	2

across phases. Richard asked more words and more so in Interventions II and III than in baseline and Intervention I.

Subjects tended to ask more words with thematic picture prompts, but on the whole differences between action and theme measures were negligible.

Holistic scoring

Table 22, the mean holistic score (scale 1-10) shows that the essays of two boys (Bevan and Aaron) were rated as slightly improved between baseline and Intervention II. However there was no difference for David's and Richard's essays between these phases. There was also little difference in the ratings given for action and theme picture prompts across phases. These differences were not satistically significant.

CONCLUSION

Experiment 2 was concerned with two major experimental areas: the effect of the nature of the stimulus prompt and of the intervention on written composition. This study aimed to answer four experimental questions which will now be discussed.

1. Did action picture prompts increase action verbs?

This experimental question had never been considered before in the literature and was tested by alternating action picture prompts with theme picture prompts.

Taking into account the finding of a significantly higher mean percentage of action verbs (total and different) for action picture prompts for all subjects compared to theme picture prompts, it can be concluded that providing subjects

with action picture prompts will stimulate more action verbs than theme prompts.

2. Were stories containing higher numbers of action verbs rated as more creative?

A consistent finding across applied behavioural studies (Ballard & Glynn, 1977; Glendinning, 1977; Maloney & Hopkins, 1973; Maloney et al., 1975) has been that the highest rating of creativity or quality, as assessed holistically, occurs when the contingency is on action verbs.

This result was also found in the present study. The increase in mean percentage total and different action verbs between baseline and Intervention II (for theme picture prompts) corresponded to an increase in the mean holistic score. This also received support from the finding that mean percentage action verbs more than doubled between baseline and Intervention III (for theme picture prompts) and the mean creativity score rose significantly between baseline and Intervention III.

However, to conclude that creativity ratings are highest when the level of action verbs are highest may be unjustified. Although action picture prompts significantly increased action verbs over theme picture prompts, these stories were not rated as more creative on the measures of mean creativity score and mean thematic maturity score. Between baseline and Intervention III, the level of action verbs for action picture prompts was the same, yet there was a significant increase in the creativity score between baseline and Intervention III for action picture prompts. This was supported statistically by the low correlation coefficient between the creativity score and different action verbs ($r = 0.38$, $r = 0.12$, $r =$

0.095, r = 0.12 for Bevan, David, Aaron and Richard respectively).

Although in some cases stories receiving the highest quality or creativity ratings had high levels of action verbs, this finding was not consistent, suggesting that other factors may also influence creativity.

It can be concluded that creativity depends on more than just action verbs alone, namely on the four global criteria used to define creativity in this thesis: originality, idea production, language usage and uniqueness of style.

3. Did action picture prompts lead to more improvements on collateral behaviours than theme picture prompts?

A review of the various ways creative writing can be stimulated, suggested that there is little conclusive evidence as to whether any one particular stimulus mode or type actually resulted in qualitative differences in written composition (Berry, 1958; Berse, 1974; Deno et al., 1980; May & Tabachnick, 1966; Sharples, 1968).

The results of this study supported the major conclusion of others studying the effect of the nature of the stimulus prompt on written composition. There was little difference whether the picture prompt was action or theme for the measures of the mechanical component (handwriting score); the productive component (number of sentences and T-units, vocabulary diversity); the conventional component (capitalization and punctuation correct, and spelling errors); the linguistic component (word usage, mature words, grammar errors, mean percentage total nouns and adjectives; mean percentage different nouns and adjectives; and T-unit length) and the contact component (words asked how to spell).

Although theme picture prompts led to higher scores than action picture prompts on the measures of mean percentage total and different adverbs and mean percentage new words (with the exception of one subject), this was offset by action picture prompts leading to higher scores than theme picture prompts on two measures, mean number of words and mean vocabulary score, for three of the four subjects.

4. The intervention's effect on written composition

The usual approach in the research literature has been to apply external teacher contingencies (such as token reinforcement, edible reinforcers and praise) to the form rather than content and ideas of written composition (e.g., Maloney & Hopkins, 1973; McKessar, 1977; Van Houten et al., 1974, 1975). In contrast to this approach Experiment 2 applied various combinations of social reinforcement and noncorrective verbal and written content feedback, delivered within a reciprocal (rather than teacher to pupil) conference context. Content and ideas rather than form were emphasized. These aspects were chosen as a result of a review of the literature which suggested that positive rather than negative feedback is more effective in improving written compositions (Clarke, 1969; Gee, 1971, 1972; Nikloff, 1966; Taylor and Hoedt, 1966); and that comments were superior to grades (Green, 1968; McKessar, 1977) and should be non-traditional and non-evaluative (Green, 1968; Kantor, 1979; Kowalski, 1983). In Experiment 2 the effects of the intervention on the target behaviour (creative responses), measures of collateral behaviours, and holistic scores were assessed.

This study demonstrated empirically that supplementing written content feedback with social reinforcement and verbal content feedback (two minutes) within a social (teacher-writer) learning context (Intervention III) was superior to providing this for only one minute (Intervention II), and to providing only written feedback from a more distant audience - the headmaster (Intervention I) and to giving no feedback or interaction at all (baseline).

Intervention III had the effect of significantly improving the content and ideas of the written compositions of four mildly retarded subjects. The mean creativity score increased for all subjects from baseline to Intervention III, and between Interventions I and II. Similarly, the mean thematic maturity score improved between baseline and Intervention III (nonsignificant), but at varying rates. Thirdly, the essays of two boys (no change for the other two) were rated holistically as slightly improved from baseline to Intervention III.

The findings also suggested that content feedback and social reinforcement provided within a social learning context, not only produced direct improvement in the content and ideas of written composition, but led to generalized improvement, as assessed by measures of collateral behaviours.

Increases (usually most noticeable between baseline and Intervention III) occurred for mean number of words for all subjects. There was a very slight increase for T-units; and the increase for the vocabulary diversity score suggested that increased word output was not achieved through repetition. This measure increased without direct intervention and contrasts with Brigham et al's (1972) study in which minimal instruction and token reinforcement contingent on different words had little effect.

In this study, three of the four subjects increased their mean vocabulary score and the fourth subject maintained his baseline level across intervention phases. This also occurred for the mean word usage score. Increases from baseline to Intervention I and baseline to Intervention III were significant. Significant increases also occurred for mean percentage total action verbs (across intervention and between Intervention II and III), and mean percentage different action verbs (across phases and between Intervention II and III, for three subjects). Maloney et al. (1975) also found a collateral increase in action verbs, but only when contingencies were placed on adverbs. Other than a significant increase between Intervention I and II, the level remained the same across phases for mean percentage different nouns. The final improvement occurred for the mean T-unit length measure for two boys (the other two remained at the same level).

Although the boys did not improve on the following measures of collateral behaviours, overall they did remain at the same level throughout the phases. The measures included the mean handwriting score, mean number of sentences, mean percentage of punctuation correct for three subjects, mean percentage spelling errors and different adverbs, mature words, total nouns and mean percentage total adverbs. These results contrasted with Maloney et al's (1975) study in which a slight increase in total and different adverbs was obtained. However this occurred when reinforcement was contingent on action verbs which was not a feature of Experiment 2.

These findings supported research (e.g., Bording et al., 1984; Poplin et al., 1980) which argued that form will not deteriorate if content is given primary importance and

focus. These findings also supported Jerram's (1985) results. She applied a similar intervention as Experiment 2 to content and ideas and this resulted not only in an increase in story quality, but also writing rate, advanced words (similar to mature words measure in Experiment 2) and spelling accuracy.

The level of responding decreased in only a few measures of collateral behaviours and in the case of mean percentage new words was statistically significant. This decrease also occurred in Experiment 1 and in Brigham et al's (1972) study. New words occurred at a reasonable rate through intervention phases, indicating that stories were still diverse in content and the subjects were not merely relying on repetition of words used in previous stories. It also indicated that picture prompts were diverse enough to stimulate new words throughout the intervention phases.

A 15% decrease occurred for two subjects on the capitalization correct measure, 50% for the third and the fourth subject remained at the same level of responding. There was a statistically significant increase in mean grammar errors per 100 words between baseline and Intervention III; and a slight decrease for three boys (moderate for the fourth) across intervention phases (statistically significant between Interventions I and II) for mean percentage total adjectives. Finally there was a slight decrease across intervention phases and between baseline and Intervention III for mean percentage different adjectives. This contrasts with Maloney and Hopkins (1973) study in which adjectives increased, which they attributed to adjectives being a similar response class to verbs. However their intervention package (giving examples before each session; written statement of contingency on blackboard;

and team competitions and token system) was completely different to the intervention in Experiment 2.

GENERAL DISCUSSION

In Experiment 1 correspondence training was applied to the target behaviour of written output in high school age slow learners. Twenty-three collateral behaviours were also measured. A changing-criterion design was used and transfer and maintenance issues were examined.

The correspondence training procedure was demonstrated to be an effective technique. Not only did it significantly increase the target behaviour, word production, for all subjects, it also generalized to numerous untrained collateral behaviours. Generally intervention gains remained at the same level or improved when transfer of control was shifted from the experimenter to the classroom teacher. The teacher found the procedure easy to implement and intended to apply correspondence training with her English class again next year. The maintenance phase results demonstrated that by merely reinforcing the verbal promise, without reinforcing its correspondence with actual performance, gains tended to remain at the same level of responding, or improved.

Risley's (1977) four essential components of the correspondence training procedure were all incorporated in Experiment 1. These were a description of the required target behaviour, guidance and reinforcement of the child's verbal statements, feedback to the audience and audience commentary on the verbal-nonverbal correspondence.

A description of the target behaviour was discussed with every student individually. This antecedent component of correspondence training involved the self-management skill of determining one's own performance standard. Each

subject was responsible for setting the number of words he felt he could reach. However he did receive guidance and reinforcement of his verbal statement from the audience, teacher or experimenter, so that he did not choose a level of words too high or too low for his ability. With respect to the consequent component of correspondence training, feedback was given to the audience as to the true performance so that accurate commentary on the correspondence relationship could be given. The true performance achieved was reported to the audience by the subject himself. Thus he was responsible for implementing another self-management skill, that of recording the number of words he actually wrote. The final essential component was that an audience commented on the verbal-nonverbal correspondence. The correspondence was reinforced in two ways. Social reinforcement in the form of verbal praise was given, since schools operate verbally, and tangible reinforcement is less appropriate with older pupils and procedurally problematic. The subjects received performance feedback when they counted the number of words they actually wrote. The audience also provided performance feedback by commenting on the correspondence.

In contrast to the usual experiences of failure these students are exposed to, they had the opportunity through the correspondence training paradigm to receive regular, frequent, consistent and immediate feedback. This occurred for a behaviour whose level they set and for which they recorded the performance level achieved to see if it corresponded. These factors tie in with Bandura's (1977) self-efficacy theory, such that the correspondence training procedure provided subjects with an opportunity to interrupt their history

of failure experiences. By progressively increasing the required level and successfully reaching it, the slow learner experienced mastery through his own efforts. Bandura (1977) has argued that performance accomplishment is the most successful way of influencing and raising a person's sense of efficacy.

The positive findings in Experiment 1 can also be related to Glynn's (1985a, 1985b, 1985c) responsive learning context. Glynn (1985a) suggested elements of this context were essential to promoting learning. Indeed, all these elements (providing the learner with an opportunity to initiate; the writer engaging in shared activity with a more skilled person with whom he shares a positive relationship; reciprocity or mutual influence; and regular responsive feedback) are very much incorporated into the corresponding training paradigm.

The findings of Experiment 1 suggest that it is possible to operationally define aspects of complex academic behaviour, such as written composition, and successfully improve such aspects in a slow learning high school population using the correspondence training paradigm. It is hoped these results will stimulate teachers to apply the correspondence training paradigm in the classroom. Its application in Experiment 1 had all six characteristics which Hopkins et al. (1971) suggested any general motivational system should have. Correspondence training was easy to manage and required minimal teacher time and training; it was based on generalized reinforcement; and its low cost allowed immediate establishment at the school. It can be applied to a variety of academic behaviours and effectively

improve the skills of various ability children (Whitman et al., 1982).

It is recommended that research into the correspondence training paradigm receive future attention. Several issues still need to be resolved, namely, why correspondence training is effective and what processes and key components are involved. There is still a need to investigate how to programme generalized verbal control across dissimilar behaviours. With respect to maintenance, research is required into the environmental variables that promote this effect. Although maintenance is currently implemented by reinforcing verbal behaviour only (rather than the correspondence) is it possible to achieve maintenance by going one step further and eliminating reinforcement of verbal behaviour as well?

In Experiment 2 feedback and social reinforcement were contingent on story content in mildly retarded boys. The effect of type of picture prompt was assessed, twenty-four collateral behaviours were measured and an alternating treatments design was used.

Experiment 2 demonstrated that although action picture prompts led to more action verbs than theme picture prompts, creativity or quality ratings were in some cases independent of the mean level of action verbs. Although the nature of the picture prompts had few differential aspects on written composition, it is recommended, considering the paucity of studies on the nature of stimulus prompts, that future research continues in this area. An example of such research is the application of an alternating treatments design to assess whether colour pictures are more stimulating to writers and lead to more interesting stories than black and white picture prompts.

More importantly this study demonstrated that interventions for written composition could successfully diverge from the usual teacher-delivered external reinforcements contingent on the form components of written composition. Experiment 2 supported Jerram's (1985) findings that social reinforcement and content feedback (verbal and written) delivered within a responsive learning context led not only to improvements in the target behaviour pertaining to content, but in addition generalized improvement or maintained levels of responding in most collateral behaviours, both qualitative and quantitative. Form did not deteriorate when content was focussed on.

These conclusions can be explained in terms of the responsive learning context (Glynn, 1985a, 1985b, 1985c) within which content feedback and social reinforcement were delivered. In baseline this context was absent and written compositions were qualitatively and quantitatively poor in standard. Although in Intervention I, the audience (headmaster) delivering written content feedback was introduced, many properties of a responsive learning context were still absent. Written compositions improved slightly in response to the introduction of this intervention. In Intervention II more properties of the responsive learning context were introduced, with the addition of verbal content feedback and social reinforcement delivered in a one-to-one conference for one minute by the classroom teacher, who shared a closer relationship with the subjects. These features were strengthened in Intervention III when the conference was extended to two minutes. For this explanation to have validity, one would expect improvements from baseline to be greatest in Intervention III. This occurred.

Intervention III exhibited all the features that Glynn (1985a, 1985b, 1985c) suggested were essential to a responsive learning context. Firstly it provided an opportunity for the writer to initiate the compositional process by giving him a picture prompt to which he was free to reply in whatever way he chose. Direct teacher intervention was reduced with the teacher not intervening during the half hour in which stories were written. There was no reiteration of the importance of form for example. Secondly, the conference preceding the writing session gave the writer an opportunity to engage in a shared activity with a more skilled person with whom he had a social relationship. Bronfenbrenner's (1979) "primary developmental context" referred to similar features inherent in this conference. His requirement of the child having a positive emotional relationship with a person under whose guidance he engaged in progressively more complex behaviours, was fulfilled. The third feature, reciprocity or mutual influence, was evident in the progressive improvement of the subject's writing as the features of the responsive learning context developed over phases. The writing was influenced by the emphasis on content and ideas, not errors, as Glynn (1985b) has prescribed. Reciprocity demands that each party modifies the behaviour of the other, and in fact, though it was less obvious, writers did modify the audience's behaviour. This was evident in the pleasure and responsive interest the audience expressed in their written content feedback to the writer. Finally, and perhaps most importantly, as the literature review revealed, the audience let the writer know that the writing affected him, through his regular, responsive, noncorrective and positive feedback.

Several issues and recommendations arise from this experiment. The first issue concerns choice. To promote initiations by the writer, it is felt that he should be offered a choice of pictures, so he finds one of interest to himself. This did not occur in Experiment 2. Future researchers could perhaps consider this option. Secondly, it has been recommended that the more immediate and specific the consequences, the more effective they are in maximizing reinforcement (e.g., Kazdin, 1984). In Experiment 2, when subjects only wrote essays every second day, they did not receive immediate reinforcement. Would the results in Experiment 2 have been stronger if reinforcement had not been delayed for a day? Future research could perhaps address this issue. The different results obtained for Intervention phases I, II and III, also raise the issue of how much and how often audience feedback should be given, to obtain a significant behavioural change. Again, future research could address this issue.

Thirdly, maintenance was neither programmed nor assessed in Experiment 2 because of time constraints. Although short-term improvements were impressive, the question of whether or not they would be maintained in the long-term needs answering.

Experiment 2 provided reinforcement and feedback on content and ideas, and in particular creative responses when they occurred. Maltzman (1960) suggested that this should foster originality, and indeed creative measures in the present study did increase. Maltzman (1960) further suggested that when creative responses do not occur or occur infrequently, one should first devise a procedure to increase their

frequency and only thereafter provide reinforcement, appraisal or recognition. Even though creative responses increased in Experiment 2, it is felt that before being reinforced for such responses, the subjects could have benefitted from a procedure to increase their low level of creative responses in the first place. This is perhaps a consideration for those interested in future research in this area.

With regard to internal validity (Barlow & Hersen, 1984), it can be concluded that the alternating treatments design ruled out the possibility that the findings in Experiment 2 were due to history, maturation, regression and changes in the measurement device, mortality, testing and other such factors (Barlow & Hayes, 1979). The biggest threat to external validity may have come from multiple-treatment interference. However the alternating treatments design minimized this possibility by providing only one treatment, theme or action, per session. Sequence effects were not controlled for in this study. It is suggested that future studies overcome this confounding effect by returning to baseline after each intervention phase, or by including a baseline control in each phase.

Written compositions were evaluated in the present study by direct assessment. From the range of scoring methods available, analytic and holistic scoring were chosen on the basis of literature review findings (e.g., Veal & Hudson, 1983).

There is controversy in the literature as to which of these two scoring procedures is superior. It is difficult to compare analytic and holistic approaches as the strength of

each rests on its own set of diverse assumptions. With respect to interscorer reliability, the research literature has consistently suggested that analytic scoring is more reliable than holistic scoring, though few comparison studies have been conducted (Cast, 1939, 1940). With respect to the reliability of holistic scoring, findings are contradictory (e.g., Brown, 1981; Diederich, 1974; Stein, 1983).

The findings of Experiments 1 and 2 suggested that analytic scoring was more reliable than holistic scoring. Analytic scoring resulted in high interscorer agreement (80-100%). The agreement was higher in Experiment 2 than Experiment 1, probably because Experiment 2 essays were shorter and less complex. Reliability percentages were highest for objective measures, and lowest for subjective ones such as creativity. High reliability levels obtained in this study can be attributed to several steps taken by the experimenter: scorers came from homogeneous backgrounds, they were trained and their performance was monitored periodically, consensus marking was adopted and scorers were provided with objective criteria and clearcut rules.

Variability in scorer assessment for holistic scoring which occurred in Experiments 1 and 2 was also found in other research (Diederich, 1974; Diederich et al., 1961; Remondino, 1959; Stein, 1983). As with analytic scoring, reliability percentages were lower for Experiment 1 than 2. Seventy-five percent of written composition samples were rated reliably in Experiment 2, as compared with 48% in Experiment 1. However in both experiments those essays that were unreliably rated, fluctuated greatly in points given by different scorers for the same essay. This is attributed to the lack of

training and clear criteria given to scorers and subjectivity involved with holistic scoring.

In comparing analytic and holistic scoring, both Veal and Hudson (1983) and Freedman (1981) found the holistic approach less time-consuming than analytic scoring. In the present study, analytic measures were found to be easy to teach, score, and administer, and were inexpensive, valid and sensitive, though time inefficient. As with Veal and Hudson (1983), analytic scoring was found to offer a more detailed analysis to the researcher. However in contrast to Freedman's (1981) claims, holistic scoring did not offer much the same information as analytic scoring. Holistic scoring provided only a global figure, pertaining to specific criteria known only to the scorer.

GENERAL RECOMMENDATIONS

Experiments 1 and 2 were limited to only one kind of writing, narrative, in only one social context, the school classroom. Future research could assess whether similar results would be obtained with a different audience, assignment or context.

Experiments 1 and 2 were limited to measuring only qualitative and quantitative changes as a result of the intervention. Unfortunately, they failed to measure changes in students' attitudes towards writing after participating in the intervention programme, a change reported in the research literature (Brigham et al., 1972; Glendinning, 1977; McKessar, 1977; Mandler & Monsen, 1985; Van Houten et al., 1974). It is recommended that attitude changes be assessed

in future research, since this change is as important as qualitative and quantitative improvements.

Both experiments dealt with a male population. Research has variously reported boys to be more affected by teacher comments than girls (Sweet, 1966), neither sex significantly affected (Gee, 1972), and females better writers than males (Martin, 1972, 1975; Meier, McCarthy & Schmeck, 1984). It is recommended that future written composition research be conducted with both sexes, so as to assess whether there are distinct differences in attitude, approach and technique, and responsiveness to intervention between the sexes, which then need to be catered for in the classroom.

Behavioural research reviewed in this study and by Kazdin (1973) assessed few if any measures of collateral behaviours. The results of Experiments 1 and 2 point to the importance of measuring these behaviours. Future studies need to give attention to this issue.

Many issues in written composition have been discussed: disagreement as to what good writing is, problematic interventions and assessment procedures. With respect to the quality of research, a great deal of literature has been philosophical or theoretical commentary with little empirical support. In future studies, subjects and settings need to be described more fully, and data on a more comprehensive range of measures, especially of collateral behaviours, need to be gathered.

An appropriate conclusion to be drawn from this study would be the need for more research. The field of written composition has scarcely been explored.

> "Complex truth is always an aggregate,
> each of us offers only part of an
> evolving mosaic".
>
> (Loban, 1976, p.90)

Agras, W.S., & Berkowitz, R. (1980). Clinical research in behaviour therapy: Halfway there? Behaviour Therapy, 11, 472-487.

Alloway, J.E. (1980). Holistic evaluation of student's writing. In Northwest regional education laboratory. Writing assessment for the 1980's (pp.85-90). Proceedings of a National Conference on the Assessment of Writing, Boulder, CO.

Anastasiow, N. (1972). Review of the picture story language test. In O.K. Buros (Ed.), Seventh mental measurement yearbook (pp.485-487). Highland Park, N J : Cryphon.

Anderson, C.C. (1960). The new STEP essay test as a measure of composition ability. Educational and Psychological Measurement, 20, 95-102.

Applebee, A.N. (1979, October). Trends in written composition. Paper presented at the Midwest School Improvement Forum, Milwaukee, WI.

Applebee, A.L. (1981). Writing in the secondary school: English and the content areas (Report No. 21). Urbana, IL: National Council of Teachers of English.

Arndt, T. (1980). Effects of timing and feedback on output and legibility in story writing. Unpublished term paper, Department of Education, University of Auckland.

Arnold, L.V. (1964). Writer's cramp and eyestrain - Are they paying off? English Journal, 53, 10-15.

Artley, S.A. (1950). Research concerning the interrelationships among the language arts. Elementary English, 27, 527-537.

Ayllon, T., Layman, D., & Burke, S. (1972). Disruptive behaviour and reinforcement of academic performance. Psychological Record, 22, 315-323.

Ayllon, T., & Roberts, M.D. (1974). Eliminating discipline problems by strengthening academic performance. Journal of Applied Behaviour Analysis, 7, 71-6.

Baer, D.M., Wolf, M.M., & Risley, T.R. (1968). Some current dimensions of applied behaviour analysis. Journal of Applied Behaviour Analysis, 1, 91-7.

Baer, R.A., Osnes, P.G. & Stokes, T.F. (1983). Training generalized correspondence between verbal behaviour at school and nonverbal behaviour at home. Education and Treatment of Children, 6, 379-388.

Baer, R.A., Williams, J.A., Osnes, P.G., & Stokes, T.F. (1984). Delayed reinforcement as an indiscriminable contingency in verbal/nonverbal correspondence training. Journal of Applied Behaviour Analysis, 17, 429-440.

Ballard, K.D. (1983). Teaching children to do what they say they will do: A review of research with suggested applications for exceptional children. Exceptional Child, 30, 119-125.

Ballard, K.D., & Glynn, T. (1975). Behavioural self-management in story writing with elementary school children. Journal of Applied Behaviour Analysis, 8, 387-398.

Ballard, K.D., & Jenner, L. (1981). Establishing correspondence between saying and doing as a procedure for increasing the social behaviours of two elementary school children. Exceptional Child, 28, 55-62.

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behaviour change. Psychological Review, 84, 191-215.

Barenbaum, E.M. (1983). Writing in the special class. Topics in Learning and Learning Disabilities, 3, 12-20.

Barlow, D.H., & Hayes, S.C. (1979). Alternating treatments design: One strategy for comparing the effects of two treatments in a single subject. Journal of Applied Behaviour Analysis, 12, 199-210.

Barlow, D.H., & Hersen, M. (1984). Single case experimental designs. Strategies for studying behaviour change (2nd (ed.). New York: Pergamon Press.

Barr, J.E., & Lambourne, R.D. (1984). Analysing spelling performance on a range of purposeful writing tasks. Educational Psychology, 4, 297-311.

Barringer, C., & Gholson, B. (1979). Effects of type and combination of feedback upon conceptual learning by children: Implications for research in academic learning. Review of Educational Research, 49, 459-478.

Barritt, L.S,, & Kroll, B.M. (1978). Some implications of cognitive-developmental psychology for research in composing. In C.R. Cooper, & L. Odell (Eds), Research on composing: Points of departure (pp.49-57). Urbana, IL: National Council of Teachers of English.

Barron, F. (1969). Creative person and creative process. New York: Holt, Rinehart and Winson.

Beach, R., & Bridwell, L.S. (Eds). (1984). New directions in composition research. New York: Guilford Press.

Bem, S.L. (1967). Verbal self-control: The establishment of effective self-instruction. Journal of Experimental Psychology, 74, 485-491.

Bennett, S.G. (1983). Park the red pencil. Educational Horizons, 61, 118-121.

Bereiter, C., & Scardamalia, M. (1983). Levels of inquiry in writing research. In P. Mosenthal, S. Walmsley, & L. Tamor (Eds). Research on writing: Principles and methods (pp.3-25). London: Longman.

Berry, E. (1958). Film and creative expression. Elementary English, 35, 383-386.

Berse, P. (1974). Criteria for the assessment of pupils' compositions. Educational Research, 17, 54-61.

Bertram, B., Collins, A., Rubin, A.D., & Gentner, D. (1982). Three perspectives on writing. Educational Psychologist, 17, 131-145.

Biberstine, R.D. (1967). A study of selected areas of teacher influence upon the written composition of fourth graders. Dissertation Abstracts International, 27, 3604A.

Bijou, S.W., Birnbrauer, J.S., Kidder, J.D., & Tague, C.E. (1966). Programmed instruction as an approach to teaching of reading, writing and arithmetic to retarded children. Psychological Record, 16, 505-522.

Birch, D. (1966). Verbal control of nonverbal behaviour. Journal of Experimental Child Psychology, 4, 266-275.

Birnbrauer, J.S., Wolf, M.M., Kidder, J.D., & Tague, C.E. (1965). Classroom behaviour of retarded pupils with token reinforcement. Journal of Experimental Child Psychology, 2, 219-235.

Bissex, G.L. (1982). Writing conferences. Alternative to the red pencil. Learning, 11, 74-77.

Blake, R.W. (1976). How to talk to a writer, or forward to fundamentals in teaching writing. English Journal, 65, 49-55.

Bloom, R.B., & Bourdon, L. (1980). Types and frequencies of teachers' written instructional feedback. Journal of Educational Research, 74, 13-15.

Bording, C., McLaughlin, T.F., & Williams, R.C. (1984). Effects of free time on grammar skills of adolescent handicapped students. Journal of Educational Research, 77, 312-318.

Borkowski, J.G., Cavanaugh, J.C., & Reichardts, G.J. (1978). Maintenance of children's rehearsal strategies: Effects of children's amount of training and strategy form. Journal of Experimental Child Psychology, 26, 288-298.

Bracewell, R.J., Frederiksen, C.H., & Frederiksen, J.D. (1982). Cognitive processes in composing and comprehending discourse. Educational Psychologist, 17, 146-164.

Braddock, R. (1969). English composition. In R.L. Ebel (Ed.), Encyclopedia of educational research (4th ed.) (pp.443-461). New York: Macmillan.

Braddock, R., Lloyd-Jones, R., & Schoer, L. (1963). Research in written composition. Urbana, IL: National Council of Teachers of English.

Braun, C., & Klassen, B. (1973). A transformational analysis of written structures of children representing varying ethno-linguistic communities. Research in the Teaching of English, 7, 312-323.

Brause, R.S. (1979, March). Cognition and written language. Analogies to spoken language. In A. Purves (Ed.), Cognition and written language: A symposium (pp.27-33). Paper presented at the Symposium on cognition and written language, Pittsburgh, PA.

Breland, H.M., & Gaynor, J.L. (1979). A comparison of direct and indirect assessments of writing skill. Journal of Educational Measurement, 16, 119-128.

Bridwell, L.S. (1980). Revising strategies in twelfth grade students' transactional writing. Research in the Teaching of English, 14, 197-222.

Briggs, D. (1970). The influence of handwriting on assessment. Educational Research, 13, 50-55.

Brigham, T.A., Burt, D., & Edwards, R. (1976). An application of operant principles to instruction. Educational Technology, 16, 45-48.

Brigham, T.A., Graubard, P.S., & Stans, A. (1972). Analysis of the effects of sequential reinforcement contingencies on aspects of composition. Journal of Applied Behaviour Analysis, 5, 421-429.

Brimner, L.D. (1982). The effects of praise and blame on writing. English Journal, 71, 58-60.

Britton, J. (1972). Writing to learn and learning to write. In The humanity of English, NCTE Distinguished Lectures. Urbana, IL: National Council of Teachers of English.

Britton, J. (1978). The composing processes and the functions of writing. In C.R. Cooper, & L. Odell (Eds), Research on composing: Points of departure. Urbana, IL: National Council of Teachers of English.

Brodsky, G. (1967). The relation between verbal and non-verbal behaviour change. Behavior Research and Therapy, 5, 183-191.

Bronfenbrenner, U. (1979). Contexts of child rearing: problems and prospectives. American Psychologist, 34, 844-850.

Broughton, S.F., & Lahey, B.B. (1978). Direct and collateral effects of positive reinforcement, response cost, and mixed contingencies for academic performance. Journal of School Psychology, 16, 126-136.

Brown, B.E. (1981). National assessments of writing ability. In C.H. Frederiksen, & J.F. Dominic (Eds), Writing: The nature, development and teaching of written communication. (Vol. 2) Writing: Process, development, and communication (pp.31-38). Hillsdale, NJ: Lawrence Erlbaum.

Brown, B.E. (1983). Reflections on recent research in composition. English Journal, 72, 51-54.

Brozick, J.R. (1978). New perspectives on compositions: A review of the literature. Journal of Aesthetic Education, 12, 83-92.

Bryant, P.E., & Bradley, C. (1983). Psychological strategies and the development of reading and writing. In M. Martlew (Ed.), The psychology of written language: Developmental and educational perspectives (pp.163-178). Chichester: John Wiley and Sons.

Burns, P.C. (1974). Diagnostic teaching of the language arts. Itasca, IL: Peacock.

Burns, P.C. (1980). Assessment and correction of language arts difficulties. Columbus, OH: Merrill.

Burron, D., & Bucher, B. (1978). Self-instructions as discriminative cues for rule-breaking or rule-following. Journal of Experimental Child Psychology, 26, 46-57.

Burton, D.L., & Arnold, L.V. (1963). Effects of frequency in writing and intensity of teacher evaluation upon high school students performance in written composition. Tallahassee, FL: Florida State University.

Buxton, E.W. (1958). An experiment to test the effects of writing frequency and guided practice upon student's skill in written expression. Doctoral dissertation, Stanford University.

Calabrese, M.E. (1982). I don't grade papers anymore. English Journal, 71, 28-31.

Campbell, J.A., & Willis, J. (1978). Modifying components of "creative behaviour" in the natural environment. Behaviour Modification, 2, 549-564.

Carlson, R.K. (1961). Seventeen qualities of original writing. Elementary English, 38, 576-579.

Carlson, R.K. (1963). Recent research in originality. Elementary English, 40, 583-589.

Carlson, R.K. (1965). An originality story scale. Elementary School Journal, 65, 366-374.

Carlson, R.K. (1973). Sparkling words: Two hundred and twenty-five practical and creative writing ideas. Geneva, IL: Paladin House.

Carnicelli, F. (1980). The writing conference: A one-to-one conversation. In T. Donovan and B. McClelland (Eds), Eight approaches to teaching composition. Urbana, IL: National Council of Teachers of English.

Carroll, J.B. (1964). Language and Thought. Englewood Cliffs, N.J. Prentice-Hall.

Cartwright, G.P. (1968). Written language abilities of educable mentally retarded and normal children. American Journal of Mental Deficiency, 72, 499-505.

Cast, B.M.D. (1939). The efficacy of different methods of marking English composition (Part I). British Journal of Educational Psychology, 9, 257-269.

Cast, B.M.D. (1940). The efficacy of different methods of marking English composition (Part II). British Journal of Educational Psychology, 10, 49-60.

Cazden, C.B. (1972). Child language and education. New York: Holt, Rinehart and Winston.

Chall, J.S., & Jacobs, V.A. (1983). Writing and reading in the elementary grades: Developmental trends among low SES children. Language Arts, 60, 617-626.

Charney, D. (1984). The validity of using holistic scoring to evaluate writing: A critical overview. Research in the Teaching of English, 18, 65-81.

Chase, C.I. (1968). The impact of some obvious variables on essay test scores. Journal of Educational Measurement, 5, 315-318.

Chomsky, C. (1971). Write first, read later. Childhood Education, 47, 296-299.

Christie, J.R., & Ballard, K.D. (1983). Teacher implementation of a say-do strategy for increasing story-writing behaviours with elementary school children. A paper presented at the New Zealand Psychological Society Annual Conference, University of Auckland.

Ciani, A.J. (1976). Syntactic maturity and vocabulary diversity. Research in the Teaching of English, 10, 150.

Cioffi, G. (1984). Observing composing behaviours of primary age children: The interaction of oral and written language. In R. Beach, & L.S. Bridwell (Eds), New directions in composition research (pp.171-190). New York: Guilford Press.

Clark, G.R. (1954). Writing situations to which children respond. Elementary English, 31, 150-155.

Clarke, G.A. (1969). Interpreting the penciled scrawl: A problem in teacher theme evaluation. MA in Teaching Thesis, University of Chicago.

Clay, M.M. (1975). What did I write? The beginnings of creative writing. Auckland, NZ: Heinemann Educational Books.

Clay, M.M. (1980). Early writing and reading: Reciprocal gains. In Clarke, M.M., & Glynn, T. (Eds), Reading and writing for the child with difficulties (pp.27-43). Education Review, Occasional Publications, No. 8, University of Birmingham.

Coffman, W.E. (1971). Essay examination. In R.L. Thorndike (Ed.), Educational measurement (pp.271-302). Washington, DC: American Council on Education.

Coffman, W.E. (1976). On the reliability of ratings of essay examinations. In W. Mehrens (Ed.), Readings in measurement and evaluation in education and psychology (pp.93-103). New York: Holt, Rinehart and Winston.

Coleman, D.R. (1983). Effects of the use of a writing scale by gifted primary students. Gifted Child Quarterly, 27, 114-121.

Combs, W. (1978). Notes and comments. Research in the Teaching of English, 12, 282-284.

Conlan, G. (1978). How the essay in the College Board English Composition test is scored. An introduction to the reading for readers. Princeton, NJ: Educational Testing Service.

Connors, R.J. (1985). Mechanical correctness as a focus in composition instruction. College Composition and Communication, 36, 61-72.

Cooper, C.R. (1975a). Measuring growth in writing. English Journal, 64, 111-119.

Cooper, C.R. (1975b). Oral and written communication. English Journal, 64, 72-74.

Cooper, C.R. (1977). Holistic evaluation of writing. In C.R. Cooper, & L. Odell (Eds), Evaluating writing: Describing, measuring, judging (pp.3-31). Urbana, IL: National Council of Teachers of English.

Cooper, C.R. (Ed.). (1981). The nature and measurement of competency in English. Urbana, IL: National Council of Teachers of English.

Cooper, C.R., & Matsuhashi, A. (1983). A theory of the writing process. In M. Martlew (Ed.), The psychology of written language: Developmental and educational perspectives. Chichester: John Wiley and Sons.

Cooper, C.R., & Odell, L. (Eds.). (1977). Evaluating writing: Describing, measuring, judging. Urbana, IL: National Council of Teachers of English.

Cooper, C.R., & Odell, L. (Eds.). (1978). Research on composing: Points of departure. Urbana, Il: National Council of Teachers of English.

Copeland, R., & Hall, R.V. (1976). Behaviour modification in the classroom. Progress in Behaviour Modification, 3, 45-78.

Cramer, R.L. (1975). The nature and nurture of creative writing. Elementary School Journal, 75, 507-512.

Croft, C. (1982). Do spelling tests measure the ability to spell? Educational and Psychological Measurement, 42, 715-723.

Crouch, K.P., Rusch, F.R., & Karlan, G.R. (1984). Competitive employment: Utilizing the correspondence teaching paradigm to enhance productivity. Education and Training of the Mentally Retarded, 7, 268-275.

Crowhurst, M., & Piche, G.L. (1979). Audience and mode of discourse effects on syntactic complexity in writing at two grade levels. Research in the Teaching of English, 13, 101-109.

Culpepper, M., & Ramsdell, R. (1982). A comparison of a multiple choice and on essay test of writing skills. Research in the Teaching of English, 16, 295-297.

Dagenais, D.J., & Beadle, K.R. (1984). Written language: When and where to begin. Topics in Language Disorders, 4, 59-85.

Daiker, D., Kerek, A., & Morenberg, M. (1978). Sentence combining and syntactic maturity in freshman English. College Composition and Communication, 19, 36-41.

Daly, J.A., & Dickson-Markman, F. (1982). Contrast effects in evaluating essays. Journal of Educational Measurement, 19, 309-316.

Davies, A. (1969). Improving writing. English in Education, 3, 63-68.

De Beaugrande, R. (1984). Text production: Toward a science of composition. Norwood, NJ: Ablex.

Dell, W.C. (1964). Creative writing in the English classroom. English Journal, 53, 500-503.

Deno, S.L., Marston, D., & Mirkin, P. (1982). Valid measurement procedures for continuous evaluation of written expression. Exceptional Children, 48, 368-370.

Deno, S.L., Mirkin, P.K., & Marston, D. (1980). Relationships among simple measures of written expression and performance on standardized achievement tests (Report No. 22). Minnesota University: Minneapolis Institute for Research on Learning Disabilities.

De Shields, S.M., Hsieh, M.K., & Frost, D. (1984). The measuring of writing skills: Some problems and a probable solution. Educational and Psychological measurement, 44, 101-112.

Diederich, P.B. (1974). Measuring growth in English. Urbana, Il: National Council of Teachers of English.

Diederich, P.B., French, J.W., & Carlton, S.T. (1961). Factors in judgments of writing ability. Princeton, NJ: Educational Testing Service.

Dilworth, C.B., & Reising, R.W. (1979). Validity in composition evaluation. Practicing what you preach. Clearing House, 53, 43-46.

Dinon, J.S. (1978). Standardized tests of writing: Multiple choices of the wrong kind. English Journal, 67, 54-56.

Dittmer, A.E. (1974). Grading compositions: Some further evidence regarding rater reliability. High School Journal, 57, 43-6.

Dixon, E.A. (1972). Syntactic indexes and student writing performance: A paper presented at NCTE - Las Vegas, 1971. Elementary English, 49, 714-716.

Douglass, B. (1984). Variation on a theme: Writing with the learning disabled adolescent. Academic Therapy, 19, 361-363.

Dusel, W.J. (1955). Some semantic implications of theme correction. English Journal, 44, 390-397.

Dyson, A.M. (1981). Oral language: The rooting system for learning to write. Language Arts, 58, 776-784.

Eachus, T. (1971). Modification of sentence writing by deaf children. American Annals of the Deaf, 116, 29-43.

Eble, K.E. (1957). Everyman's handbook of final comments on freshman themes. College English, 19, 126-127.

Ebel, R.L., & Damrin, D.E. (1960). Tests and examinations. In C. Harris (Ed.), Encyclopaedia of educational research (pp.1502-1517). New York: Macmillan.

Edmaiston, R., & Larson, C. (1983). An investigation of oral language and writing proficiency in 3rd-grade students. Psychology in the Schools, 20, 380-387.

Edmund, N. (1958a). Relationships between prior experiences and the creative quality of stories of fifth grade children. Elementary English, 35, 248-249.

Edmund, N. (1958b). A study of the relationship between prior experience and the quality of creative writing done by seventh-grade pupils. Journal of Educational Research, 51, 481-492.

Edmund, N.R. (1959). Writing in the intermediate grades. Elementary English, 36, 491-501.

Emig, J.A. (1971). The composing processes of twelfth graders (Report No. 13). Urbana, Il: National Council of Teachers of English.

Emig, J. (1977). Writing as a mode of learning. College Composition and Communication, 28, 122-128.

Emig, J. (1978). Hand, eye, brain: Some "basics" in the writing process. In C.R. Cooper, & L. Odell (Eds), Research on composing: Points of departure (pp.59-71). Urbana, IL: National Council of Teachers of English.

Essex-Sorlie, D. (1984). Test of written language. In D.J. Keyser, & R.C. Sweetland (Eds), Test critiques (Vol. I) (pp.689-706). Kansas City, MO: Test Corporation of America.

Evans, R. (1967). A glove thrown down. Elementary English, 44, 523-527.

Evertts, E.L. (1966). Dinosaurs, witches, and anti-aircraft: Primary composition. Elementary English, 43, 109-114.

Fagerlie, A.M. (1975). Brand names stimulate storywriting. Language Arts, 52, 1017-1018.

Faigley, L.L. (1977). Measuring growth in college writing. Paper presented at the annual meeting of the conference on college composition and communication. Kansas City, MO.

Fauke, J., Burnett, J., Powers, M.A., & Sulzer-Azaroff, B. (1973). Improvement of handwriting and letter recognition skills: A behaviour modification procedure. Journal of Learning Disabilities, 6, 296-300.

Ferritor, D.E., Buckholdt, D., Hamblin, R.L., & Smith, L. (1972). The noneffects of contingent reinforcement for attending behaviour on work accomplished. Journal of Applied Behaviour Analysis, 5, 7-17.

Fillion, B., Smith, F., & Swain, M. (1976). Language "basics" for language teachers: Towards a set of universal considerations. Language Arts, 53, 740-745.

Finn, P.J. (1977). Computer-aided description of mature word choices in writing. In C.R. Cooper, & L. Odell (Eds), Evaluating writing: Describing, measuring, judging (pp.69-89). Urbana, Il: National Council of Teachers of English.

Florio-Ruane, S. (1983). What's so hard about writing? The issues for teachers and students. Elementary School Journal, 84, 93-99.

Flower, L., & Hayes, J.R. (1981). A cognitive process theory of writing. College Composition and Communication, 32, 365-387.

Follman, J.C., & Anderson, J.A. (1967). An investigation of the reliability of five procedures for grading English themes. Research in the Teaching of English, 1, 190-200.

Ford, B.G., & Renzulli, J.S. (1976). Developing the creative potential of educable mentally retarded students. Journal of Creative Behaviour, 10, 210-218.

Fowler, S., & Baer, D.M. (1981). "Do I have to be good all day?" The timing of delayed reinforcement as a factor in generalization. Journal of Applied Behaviour Analysis, 14, 13-24.

Foxx, R.M., & Jones, J.R. (1978). A remediation program for increasing the spelling achievement of elementary and junior high students. Behaviour Modification, 2, 211-230.

Frederiksen, C.H., & Dominic, J.F. (Eds). (1981). Writing: The nature, development and teaching of written communication. (Vol. 2) Writing: Process, development and communication. Hillsdale, NJ: Lawrence Erlbaum.

Freedman, S.W. (1979a). How characteristics of student essays influence teachers' evaluations. Journal of Educational Psychology, 71, 328-338.

Freedman, S.W. (1979b). Why do teachers give the grades they do? College Composition and Communication, 30, 161-164.

Freedman, S. (1981). Influences on the evaluators of expository essays: Beyond the text. Research in the Teaching of English, 15, 245-255.

Freedman, S.W. (1982). Language assessment and writing disorders. Topics in Language Disorders, 2, 34-44.

Freedman, S., & Calfee, R. (1983). Holistic assessment of writing: Experimental design and cognitive theory. In P. Mosenthal, L. Tamor, & S. Walmsley (Eds), Research in writing: Principles and methods. New York: Longman.

Fry, E. (1977). Elementary reading instruction. New York: McGraw-Hill.

Funk, H.D. (1975). Is it worth the time and effort? Language Arts, 52, 1015-1016.

Gage, N.L. (1964). This side of paradigms: The state of research on teaching English. In D.H. Russell, M.J. Early, E. Farrell, Research design and the teaching of English: Proceedings of the San Francisco Conference (pp.22-31). Champaign, IL: National Council of Teachers of English.

Galda, L. (1984). The relations between reading and writing in young children. In R. Beach, & L.S. Bridwell (Eds), New directions in composition research (pp.191-204). New York: Guilford Press.

Gee, T.C. (1971). The effects of written comment on expository composition. Dissertation Abstracts International, 31, 3412A.

Gee, T.C. (1972). Students' responses to teacher comments. Research in the Teaching of English, 6, 212-221.

Gere, A.R. (1980). Written composition: Toward a theory of evaluation. College English, 42, 44-48.

Getzels, J.W., & Jackson, P.W. (1960). The study of giftedness: A multi-dimensional approach. The Gifted Student. Cooperation Research Monograph No. 2. U.S. Dept. of Health, Education, and Welfare. Washington, DC: U.S. Government Printing Office.

Getzels, J.W., & Jackson, P.W. (1962). Creativity and intelligence. Explorations with gifted students. New York: John Wiley and Sons.

Glendinning, S. (1977). Reinforcement of attentiveness and academic performance in an adjustment class. Unpublished master's thesis, Department of Education, University of Auckland.

Glover, J.A. (1979). The effectiveness of reinforcement and practice for enhancing the creative writing of elementary school children. Journal of Applied Behaviour Analysis, 12, 487.

Glover, J., & Gary, A.L. (1976). Procedures to increase some aspects of creativity. Journal of Applied Behaviour Analysis, 9, 79-84.

Glynn, E.L. (1970). Classroom applications of self-determined reinforcement. Journal of Applied Behaviour Analysis, 3, 123-132.

Glynn, T. (1981). Behavioural research in remedial education. More power to the parents. In K. Wheldall (Ed.), The behaviourist in the classroom: Aspects of applied behaviour analysis in British educational contexts. Educational Review Offset Publications, No. 1, University of Birmingham.

Glynn, T. (1982a). A behavioural approach to teaching. In K. Wheldall, & R. Riding (Eds), Psychological aspects of learning and teaching. London: Croom Helm.

Glynn, T. (1982b). Antecedent control of behaviour in educational contexts. Educational Psychology, 2, 215-229.

Glynn, T. (1985a). Contexts for independent learning. Educational Psychology, 5, 5-15.

Glynn, T. (1985b, January). Writing as an interactive social process. Invited address. Fourth NZEAS Conference

Glynn, T. (1985c, September). Interactive contexts for acquiring literacy skills. Invited address. First annual conference. Queensland Guidance Officers Association, Surfers Paradise, Queensland.

Glynn, E.L., McNaughton, S.S., & Wotherspoon, A.T. (1974). Modification of reading, writing and attending behaviour in a special class for retarded children. Unpublished paper, Department of Education, University of Auckland.

Glynn, E.L., Wotherspoon, A.T., & Harbridge, R. (1976). Towards self-management of generative writing in the special class. Exceptional Child, 23, 31-45.

Godstalk, F.I., Swineford, F., & Coffman, W.E. (1966). The measurement of writing ability. New York: College Entrance Examination Board.

Goetz, E.M., & Baer, D.M. (1971). Descriptive social reinforcement of creative blockbuilding by young children. In E.A. Ramp, & B.L. Hopkins (Eds), A new direction for education: Behaviour analysis (Vol. 1) (pp.72-79). Laurence, KS: University of Kansas.

Goetz, E.M., & Baer, D.M. (1973). Social control of form diversity and the emergence of new forms in children's blockbuilding. Journal of Applied Behaviour Analysis, 6, 209-217.

Goetz, E.M., & Salmonson, M.M. (1972). The effect of general and descriptive reinforcement on creativity in easel painting. In G.B. Semb (Ed.), Behaviour analysis in education (pp.53-61). Laurence, KS: University of Kansas Press.

Golden, J.M. (1980). The writer's side: Writing for a purpose and an audience. Language Arts, 57, 756-762.

Golub, L.S., & Frederick, W.C. (1970). An analysis of children's writing under different stimulus conditions. Research in the Teaching of English, 4, 168-180.

Gowan, J.C., Demos, G.D., & Torrance, E.P. (1967). Creativity: Its educational implications. New York: John Wiley and Sons.

Graham, S. (1982). Composition research and practice: A unified approach. Focus on Exceptional Children, 14(8), 1-16.

Grandholm, S. (1983). Ask with a picture: Answer with a story. Learning, 11, 126.

Graves, D.H. (1975). An examination of the writing processes of seven year old children. Research in the Teaching of English, 9, 227-241.

Graves, D.H. (1978a). Balance the basics: Let them write. Learning, 6, 30-33.

Graves, D.H. (1978b). We won't let them write. Language Arts, 55, 635-640.

Graves, D.H. (1980). A new look at writing research. Language Arts, 57, 913-919.

Graves, D.H. (1981). Research update: Writing research for the 80's: What is needed. Language Arts, 58, 197-206.

Graves, D.H. (1983). Writing: Teachers and children at work. London: Heinemann Educational Books.

Green, J.L. (1968). A method for writing comments on student themes. English Journal, 57, 215-220.

Green, G.M., & Morgan, J.L. (1981). Writing ability as a function of the appreciation of differences between oral and written communication. In C.M. Frederiksen, & J.F. Dominic (Eds), Writing: The nature, development and teaching of written communication. (Vol. 2) Writing: Process, development and communication (pp.177-188). Hillsdale, NJ: Lawrence Erlbaum.

Greenhalgh, C., & Townsend, D. (1981). Evaluating students' writing holistically - An alternative approach. Language Arts, 58, 811-822.

Gripps, C., & Ewen, E. (1974). Scoring written work in English as a second language: The use of the T-unit. Educational Research, 16, 121-125.

Grobe, C. (1981). Syntactic maturity, mechanics and vocabulary predictors of quality ratings. Research in the Teaching of English, 15, 75-85.

Groff, P. (1978). Children's oral language and their written composition. Elementary School Journal, 78, 181-191.

Grossman, H.J. (1983). Classification in mental retardation. Washington, DC: American Association of Mental Deficiency.

Guilford, J.P. (1950). Creativity. American Psychologist, 5, 444-454.

Gutman, H. (1961). The biological roots of creativity. Genetic Psychological Monographs, 64, 419-458.

Hagaman, J. (1980). Toward more effective teacher response to student writing. Paper presented at the conference on college composition and communication, Washington, DC.

Hagstrum, J.H. (1964). Research in written composition. College English, 26, 53-56.

Hairston, M. (1982) The winds of change: Thomas Kuhn and the revolution in the teaching of writing. College composition and communication, 33, 76-88.

Hales, L.W., & Tokar, E. (1975). The effect of the quality of preceding responses on the grades assigned to subsequent responses to an essay question. Journal of Educational Measurement, 12, 115-117.

Hammill, D.D., & Bartel, N.R. (Eds). (1978). Teaching children with learning and behaviour problems (2nd ed.). Boston, MA: Allyn and Bacon.

Hammill, D.D., & Larsen, S.C. (1983). The test of written language. Austin, TX: PRO-ED.

Hammill, D.D., & Poplin, M. (1982). Problems in written composition. In D.D. Hammill, & N.R. Bartel (Eds), Teaching children with learning and behaviour problems (pp.93-120). Boston, MA: Allyn and Bacon.

Hansen, C.L. (1978). Writing skills. In N.G. Haring, T.C. Lovitt, M.D. Eaton, & C.L. Hansen (Eds), The fourth R: Research in the classroom (pp.93-126). Columbus, OH: Charles E. Merrill.

Haring, N.G., Lovitt, T.C., Eaton, M.D., & Hansen, C.L. (Eds). (1978). The fourth R: Research in the classroom. Columbus, OH: Charles E. Merrill.

Harris, W.H. (1977). Teacher response to student writing: A study of the response patterns of high school English teachers to determine the basis for teacher judgment of student writing. Research in the Teaching of English, 11, 175-185.

Harris, K.R., & Graham, S.E. (1985). Improving learning disabled students' composition skills: Self-control strategy training. Learning Disability Quarterly, 8, 27-36.

Harris, V.W., & Sherman, J.A. (1974). Homework assignments, consequences and classroom performance in social studies and mathematics. Journal of Applied Behaviour Analysis, 7, 505-519.

Harrop, A. (1978). Behaviour modification in the ordinary school setting. Association of Educational Psychologists Journal, 4(7), 3-15.

Harrop, A., & McCann, C. (1984). Modifying "creative writing" in the classroom. British Journal of Educational Psychology, 54, 62-72.

Hartmann, D.P., & Hall, V.R. (1976). The changing-criterion design. Journal of Applied Behaviour Analysis, 9, 527-532.

Hartog, P. Sir, & Rhodes, E.C. (1936). The marks of examiners. New York: Macmillan.

Hay, W.M., Hay, L.R., & Nelson, R.O. (1977). Direct and collateral changes in on-task and academic behaviour resulting from on-task versus academic contingencies. Behaviour Therapy, 8, 431-441.

Hayes, S.C., Rincover, A., & Solnick, J.V. (1980). The technical drift of applied behaviour analysis. Journal of Applied Behaviour Analysis, 13, 275-285.

Henning, G.H. (1984). Language test development. Rowley, MA: Newbury House.

Heward, W.L., & Eachus, H.T. (1979). Acquisition of adjectives and adverbs in sentences written by hearing impaired and aphasic children. Journal of Applied Behaviour Analysis, 12, 391-400.

Hillerich, R.L. (1971). Evaluation of written language. Elementary English, 48, 839-842.

Hillerich, R.L. (1979). Developing written expression: How to raise - not raze - writers. Language Arts, 56, 769-777.

Hirsch, E.D., Jr. (1977). The philosophy of composition. Chicago: University of Chicago Press.

Hogan, T.P., & Mishler, C. (1980). Relationships between essay tests and objective tests of language skills for elementary school students. Journal of Educational Measurement, 17, 219-227.

Holman, J., Goetz, E.M., & Baer, D.M. (1977). The training of creativity as an operant and an examination of its generalization characteristics. In B.C. Etzel, J.M. Le Blanc, & D.M. Baer (Eds), New developments in behavioural research. Theory, method and application. In honour of Sidney W. Bijou (pp.441-471). Hillsdale, NJ: Lawrence Erlbaum Associates.

Hopkins, B.L., Shutte, R.C., & Garton, K.L. (1971). The effects of access to a playroom on the rate and quality of printing and writing of first and second-grade students. Journal of Applied Behaviour Analysis, 4, 77-87.

Hops, H., Walker, H.H., & Greenwood, C.R. (1979). PEERS: A program for remediating social withdrawal in school. In L.A. Hamerlynck (Ed.), Behavioural systems for the developmentally disabled: 1. School and family environments (pp.48-86). New York: Brunner-Mazel.

Hotopf, N. (1980). Slips of the pen. In U. Frith (Ed.), Cognitive processes in spelling (pp.287-307). London: Academic Press.

Howell, M. (1956). Differentiating variables in compositions of seven-year olds. Elementary School Journal, 57, 145-149.

Huck, S.W., & Bounds, W.G. (1972). Essay grades: An interaction between graders' handwriting clarity and the neatness of examination papers. American Educational Research Journal, 9, 279-283.

Huddleston, E. (1954). Measurement of writing ability at the college entrance level: Objective vs. subjective techniques. Journal of Experimental Education, 22, 165-213.

Hughes, D.C., & Keeling, B. (1984). The use of model essays in an attempt to reduce context effects in essay scoring. Journal of Educational Measurement, 21, 277-281.

Hughes, D.C., Keeling, B., & Tuck, B.F. (1980a). The influence of context position and scoring method in essay scoring. Journal of Educational Measurement, 17, 131-135.

Hughes, D.C., Keeling, B., & Tuck, B.F. (1980b). Essay marking and the context problem. Educational Research, 22, 147-148.

Hughes, D.C., Keeling, B., & Tuck, B.F. (1983). The effects of instructions to scorers intended to reduce context effects in essay scoring. Educational and Psychological Measurement, 43, 1047-1050.

Humes, A. (1983). Research on the composing process. Review of Educational Research, 53, 201-216.

Hundert, J., Butcher, B., & Henderson, M. (1976). Increasing appropriate classroom behaviour and academic performance by reinforcing correct work alone. Psychology in the Schools, 2, 195-200.

Hunt, K.W. (1965). Grammatical structures written at three grade levels (Report No. 3). Champaign, IL: National Council of Teachers of English.

Hunt, K.W. (1966). Recent measures in syntactic development. Elementary English, 42, 732-739.

Hunt, K.W. (1970). Syntactic maturity in school children and adults. Monographs of the Society for Research in Child Development, 35, Serial No. 134.

Hunt, K.W. (1977). Early blooming and late blooming syntactic structures. In C.R. Cooper, & L. Odell (Eds), Evaluating writing: Describing, measuring, judging (pp.91-104). Urbana, IL: National Council of Teachers of English.

Irmscher, W.F. (1979). Teaching expository writing. New York: Holt, Rinehart and Winston.

Israel, A.C. (1973). Developing correspondence between verbal and nonverbal behaviour: Switching sequences. Psychological Reports, 32, 1111-1117.

Israel, A.C. (1978). Some thoughts on correspondence between saying and doing. Journal of Applied Behaviour Analysis, 11, 271-276.

Israel, A.C., & Brown, M.S. (1977). Correspondence training, prior verbal training, and control of nonverbal behaviour via control of verbal behaviour. Journal of Applied Behaviour Analysis, 10, 333-338.

Israel, A.C., & O'Leary, K.D. (1973). Developing correspondence between children's words and deeds. Child Development, 44, 575-581.

Jacobs, H.L., Zingraf, S.A., Wormuth, D.R., Hartfiel, V.F., & Hughey, J.B. (1981). Testing ESL composition: A practical approach. Rowley, MA: Newbury House.

Jacobs, L.B. (1984). Portfolio on writing (symposium). Early Years, 14, 59-72.

Jerábek, R., & Dieterich, D. (1975). Composition evaluation: The state of the art. College Composition and Communication, 26, 183-186.

Jerram, H. (1985). Written content feedback in children's story writing. Master's thesis, Department of Education, University of Auckland.

Jewett, J., & Clark, H.B. (1979). Teaching preschoolers to use appropriate dinnertime conversation: An analysis of generalization from school to home. Behaviour Therapy, 10, 589-605.

Johnson, L.V. (1962). Problems in evaluation of writing. Elementary English, 39, 570-573.

Johnson, L.V. (1967). Children's writing in three forms of composition. Elementary Education, 44, 265-269.

Johnson, W. (1944). Studies in language behaviour: I. A program of research. Psychological Monographs, 56, 1-15.

Jones, T.P. (1972). Creative learning in perspective. London: University of London Press.

Judy, S.N. (1973). Writing for the here and now: An approach to assessing student writing. English Journal, 62, 69-79.

Kaczmarek, C.M. (1980). Scoring and rating essay talks. In J.W. Oller, & K. Perkins (Eds), Research in language testing (pp.151-159). Rowley, MA: Newbury House.

Kagan, J. (Ed.). (1967). Creativity and learning. Boston, MA: Houghton Mifflin.

Kanfer, F.H., & Karoly, P. (1972). Self-control: A behaviouristic excursion into the lion's den. Behaviour Therapy, 3, 398-416.

Kantor, K.J. (1979). Appreciating children's writing. Language Arts, 56, 742-746.

Kantor, K., & Perron, J. (1977). Thinking and writing: Creativity in three modes of discourse. Language Arts, 54, 742-749.

Karlan, G.R., & Rusch, F.R. (1982). Correspondence between saying and doing: Some thoughts on defining correspondence and future directions for application. Journal of Applied Behaviour Analysis, 15, 151-162.

Karoly, P., & Dirks, M.J. (1977). Developing self-control in preschool children through correspondence training. Behaviour Therapy, 8, 398-405.

Kazdin, A.E. (1973). Methodological assessment considerations in evaluating reinforcement programs in applied settings. Journal of Applied Behaviour Analysis, 6, 517-531.

Kazdin, A.E. (1981). Behaviour modification in education: Contributions and limitations. Developmental Review, 1, 34-57.

Kazdin, A.E. (1982). Single-case research designs. Methods for clinical and applied settings. New York: Oxford University Press.

Kazdin, A.E. (1984). Behaviour modification in applied settings (3rd Ed.). Homewood, IL: Dorsey Press.

Kehl, D.G. (1970). The art of writing evaluative comments on student themes. English Journal, 59, 972-980.

Kerr, M.M., & Lambert, D.L. (1982). Behaviour modification of children's written language. Progress in Behaviour Modification, 13, 79-108.

Kincaid, G.L. (1953). Some factors affecting variations in the quality of students' writing. Unpublished doctoral dissertation, Michigan State College.

King, M.L. (1978). Research in composition: A need for theory. Research in the Teaching of English, 12, 193-202.

Kinnick, B.J. (1960). Creative writing. Todays Education, 49, 20-22.

Kirby, F.D., & Shields, F. (1972). Modification of arithmetic response rate and attending behaviour in a seventh-grade student. Journal of Applied Behaviour Analysis, 5, 79-84.

Kline, C. (1976). I know you know what I said. College English, 37, 661-662.

Koch, R. (1982). Syllogisms and superstitions: The current state of responding to writing. Language Arts, 59, 464-471.

Kowalski, J. (1983). Writing composition comments can be fun. Clearing House, 57, 39.

Kraetsch, G.A. (1981). The effects of oral instructions and training on the expansion of written language. Learning Disability Quarterly, 4, 82-90.

Kuhn, T. (1970). The structure of scientific revolutions. Chicago: University of Chicago Press.

Kulhavy, R.W. (1977). Feedback in written instruction. Review of Educational Research, 47, 211-232.

Kuykendall, C. (1975). Grammar and composition: Myths and realities. English Journal, 64, 6-7.

Lahey, B.B. (1976). Behaviour modification with learning disabilities and related problems. Progress in Behaviour Modification, 3, 173-205.

Lazdowski, W.P. (1976). Determining reading grade levels from analysis of written compositions. Dissertation Abstracts International, 37, 1504-1505A.

Lerner, J.W. (1981). Learning disabilities: Theories, diagnosis, and teaching strategies (3rd ed.). Boston: Houghton Mifflin.

Lickteig, M.J. (1981). Research-based recommendations for teachers of writing. Language Arts, 58, 44-50.

Littwin, M. (1935). An experimental investigation of the effect of presentation upon the imaginative quality of descriptive writing among elementary-school pupils. Journal of Exceptional Education, 4, 44-49.

Lloyd-Jones, R. (1970). Theoretical problems in studying creativity and composition. College Composition and Communication, 21, 261-6.

Lloyd-Jones, R. (1977). Primary trait scoring. In C.R. Cooper, & L. Odell (Eds), Evaluating writing: Describing, measuring, judging (pp.33-66). Urbana, IL: National Council of Teachers of English.

Loban, W. (1963). The language of elementary schoolchildren. Champaign, IL: National Council of Teachers of English.

Loban, W.D. (1976). Language development: Kindergarten through grade twelve (Report No. 18). Urbana, IL: National Council of Teachers of English.

Lovaas, O.I. (1961). Interaction between verbal and non-verbal behaviour. Child Development, 32, 329-336.

Lovaas, O.I. (1964). Cue properties of words: The control of operant responding by rate and content of verbal operants. Child Development, 35, 245-246.

Lovitt, T.C. (1976). Applied behaviour analysis techniques and curriculum research: Implications for instruction. In N.G. Haring, & Schiefelbusch, R.L. (Eds), Teaching special children (pp.112-156). New York: McGraw-Hill.

Lundsteen, S.W. (1976). Children learn to communicate: Language arts through creative problem solving. Englewood Cliffs, NJ: Prentice-Hall.

Luria, A.R. (1959). The directive function of speech in development. Word, 15, 341-352.

Luria, A.R. (1961). The role of speech in the regulation of normal and abnormal behaviour. New York: Liveright.

Lysakowski, R.S., & Walberg, H.J. (1981). Classroom reinforcement and learning: A quantitative synthesis. Journal of Educational Research, 75, 69-77.

McColly, W. (1970). What does educational research say about the judging of writing ability? Journal of Educational Research, 64, 148-156.

McKessar, C.J. (1977). A comparison of feedback types of written language in the classroom. Unpublished master's thesis, University of Waikato.

McLaughlin, T.F., & Malaby, J.E. (1972). Intrinsic reinforcers in a classroom token economy. Journal of Applied Behaviour Analysis, 5, 263-270.

McLaughlin, T.F., & Malaby, J.E. (1974). The utilization of an individual contingency program to control assignment completion in a token classroom: A case study. Psychology in the Schools, 11, 191-194.

McLoughlin, J.A. & Lewis, R.B. (1981). Assessing special students: Strategies and procedures. Columbus, OH: Merrill.

Macmillan, D.L., & Forness, S.R. (1970). Behaviour modifications: Limitations and liabilities. Exceptional Children, 37, 291-297.

Malgady, R.G., & Barcher, P.R. (1977). Psychological scaling of essay creativity: Effects of productivity and novelty. Journal of Educational Psychology, 69, 512-518.

Maloney, K.B., & Hopkins, B.L. (1973). The modification of sentence structure and its relationship to subjective judgements of creativity in writing. Journal of Applied Behaviour Analysis, 6, 425-433.

Maloney, K.B., Jacobson, L., & Hopkins, B.L. (1975). An analysis of the effects of lectures, requests, teacher praise and free time on the creative writing behaviours of third grade children. In E. Ramp, & G. Semb (Eds), Behaviour analysis - Areas of research and application. Englewood Cliffs, NJ: Prentice-Hall.

Maltzman, I. (1960). On the training of originality. Psychological Review, 67, 229-242.

Mander, R., & Monsen, J. (1985). The use of cognitive rehearsal strategies for increasing story writing behaviours in children: A pilot study. Paper presented at the Eighth Annual Conference of the Australian Behaviour Modification Association, Melbourne, Australia.

Marholin, D., Siegel, L.J., & Philips, D. (1976). Treatment and transfer: A search for empirical procedures. Progress in Behaviour Modification, 3, 293-342.

Marholin, D., & Steinman, W.M. (1977). Stimulus control in the classroom as a function of the behaviour reinforced. Journal of Applied Behaviour Analysis, 10, 465-478.

Markham, L.R. (1976). Influences of handwriting quality on teacher evaluation of written work. American Educational Research Journal, 13, 277-283.

Marshall, B. (1971). A survey and analysis of teachers' markings on selected compositions of average students in grades 10 and 12. Unpublished doctoral dissertation, State University of New York, Buffalo.

Marshall, J.C. (1972). Writing neatness, composition errors, and essay grades reexamined. Journal of Educational Research, 65, 213-215.

Marshall, J.C., & Powers, J.M. (1969). Writing neatness, composition errors, and essay grades. Journal of Educational Measurement, 6, 97-101.

Marshall, M. (1960). Helping children to write better themes. Peabody Journal of Education, 38, 96-99.

Martin, H.C. (1961). Testing for elegance. In Proceedings of the invitational conference on testing problems (pp. 87-96). Princeton, NJ: Educational Testing Service.

Martin, W.D. (1972). The sex factor in grading composition. Research in the Teaching of English, 6, 36-47.

Martin, W.D. (1975). Measuring children's story writing - II. Language Arts, 52, 1023-1025.

Martlew, M. (Ed.). (1983). The psychology of written language: Developmental and educational perspectives. Chichester: John Wiley and Sons.

Masterson, J.F., Earls, H., & Taylor, D. (1981). The effects of instructions and reinforcement on the handwriting of a 13-year-old child in a remedial school. Behavioural Psychotherapy, 9, 167-176.

May, F., & Tabachnick, B. (1966). Three stimuli for creative writing. Elementary School Journal, 67, 88-94.

Meckel, H.C. (1967). Research on teaching composition and literature. In N.L. Gage (Ed.), Handbook of research on teaching (pp.966-1006). Chicago: Rand McNally.

Meichenbaum, D., & Goodman, J. (1969a). The developmental control of operant motor responding by verbal operants. Journal of Experimental Child Psychology, 7, 553-565.

Meichenbaum, D., & Goodman, J. (1969). Reflection-impulsivity and verbal control of motor behaviour. Child Development, 40, 785-797.

Meichenbaum, D., & Goodman, J. (1971). Training impulsive children to talk to themselves: A means of developing self-control. Journal of Abnormal Psychology, 77, 115-126.

Meier, S., McCarthy, P.R., & Schmeck, R.R. (1984). Validity of self-efficacy as a predictor of writing performance. Cognitive Therapy and Research, 8, 107-120.

Mellon, J. (1969). Transformational sentence-combining: A method for enhancing the development of syntactic fluency in English composition (Report No. 10). Urbana, IL: National Council of Teachers of English.

Mellon, J. (1975). National assessment and the teaching of English. Urbana, IL: National Council of Teachers of English.

Meredith, V.H., & Williams, P.L. (1984). Issues in direct writing assessment: Problem identification and control. Educational Measurement: Issues and Practice, 3, 11-15.

Merrett, F.E. (1981). Studies in behaviour modification in British educational settings. Educational Psychology, 1, 13-38.

Merrett, F.E., & Wheldall, K. (1978). Playing the game: A behavioural approach to classroom management in the junior school. Educational Review, 30, 41-50.

Miller, L.K., & Schneider, R. (1970). The use of a token system in project headstart. Journal of Applied Behaviour Analysis, 3, 213-220.

Moffat, J. (1968). The structure of English. Christchurch, NZ: Pegasus Press.

Moffett, J. (1968). Teaching the universe of discourse. New York: Houghton Mifflin.

Monahan, J., & O'Leary, K.D. (1971). The effects of self-instruction on rule-breaking behaviour. Psychological Reports, 29, 1059-1066.

Moran, M. (1981). Performance of learning disabled and low achieving secondary students on formal features of a paragraph-writing test. Learning Disability Quarterly, 4, 271-280.

Morris, N.T., & Crump, W.D. (1982). Syntactic and vocabulary development in the written language of learning disabled and non-learning disabled students at four age levels. Learning Disability Quarterly, 5, 163-172.

Mosenthal, P. (1982). Toward a paradigm of children's classroom competence. Advances in Reading/Language Research, 1, 125-154.

Moslemi, M.M. (1975). The grading of creative writing essays. Research in the Teaching of English, 9, 154-161.

Moss, P.A., Cole, N.S., & Khampalikit, C. (1982). A comparison of procedures to assess written language skills in grades 10, 7, and 4. Journal of Educational Measurement, 19, 37-47.

Murray, D.M. (1984). Facets. The most important development in the last five years for high school teachers of composition (symposium). English Journal, 73, 20-23.

Myers, A., McConville, C.B., & Coffman, W.E. (1966). Simplex structure in the grading of essay tests. Educational and Psychological Measurement, 26, 41-54.

Myers, M. (1980). A procedure for writing assessment and holistic scoring. Urbana, IL: National Council of Teachers of English.

Myklebust, H.R. (1965). Development and disorders of written language (Vol. 1). Picture story Language Test. New York: Grune and Stratton.

Myklebust, H.R. (1973). Development and disorders of written language (Vol. 2). Studies of normal and exceptional children. New York: Grune and Stratton.

National Council of Teachers of English. (1983). Reading and writing. Language Arts, 60, 545-681.

Neilson, L., & Piche, G. (1981). The influence of headed nominal complexity and lexical choice on teachers' evaluation of writing. Research in the Teaching of English, 15, 65-74.

Nelson, L. (1965). Inquiry - into the influence of the assigned topic on written language. California Journal of Educational Research, 16, 100-107.

Ney, J.W. (1966). Review of grammatical structures written at three grade levels, Kellog Hunt. Language Learning, 16, 230-235.

Nichols, S. (1970). Pupil motivation: A rewarding experience. Maryland English Journal, 8, 36-41.

Nikoloff, S.E.B. (1966). The relationship of teacher standards to the written expression of fifth and sixth grade children. Dissertation Abstracts International, 26, 6560A.

Nold, E.W. (1981). Revising. In C.H. Frederiksen, & J.F. Dominic (Eds), Writing: The nature, development, and teaching of written communication. (Vol. 2) Writing: Process, development, and communication (pp.67-79). Hillsdale, NJ: Lawrence Erlbaum.

Nold, E., & Freedman, S. (1977). An analysis of readers' responses to essays. Research in the Teaching of English, 11, 164-174.

O'Dea, P. (1965). Five myths in the teaching of composition. English Journal, 54, 328-330.

Odell, L. (1979, March). Procedures for exploring writers' cognitive maturity. In A. Purves (Ed.), Cognition and written language: A symposium (pp.105-111). Paper presented at the symposium on cognition and written language, Pittsburg, PA.

Odell, L. (1981). Defining and assessing competence in writing. In C. Cooper (Ed.), The nature and measurement of competency in English (pp.95-138). Urbana, IL: National Council of Teachers of English.

Odell, L., & Cooper, C.R. (1980). Procedures for evaluating writing: Assumptions and needed research. College English, 42, 35-43.

Odell, L., Cooper, C.R., & Courts, C. (1978). Discourse theory: Implications for research in composing. In C.R. Cooper, & L. Odell (Eds), Research on composing: Points of departure (pp.1-12). Urbana, IL: National Council of Teachers of English.

O'Donnell, R.C. (1976). A critique of some indices of syntactic maturity. Research in the Teaching of English, 10, 31-38.

O'Donnell, R.C., Griffin, W.J., & Norris, R.C. (1967). Syntax of kindergarten and elementary school children: A transformational analysis (Report No. 8). Champaign, IL: National Council of Teachers of English.

O'Hare, F. (1973). Sentence combining: Improving student writing without formal grammar instruction (Report No. 15). Urbana, IL: National Council of Teachers of English.

O'Leary, K.D. (1968). The effects of self-instruction on immoral behaviour. Journal of Experimental Child Psychology, 6, 297-301.

O'Leary, K.D., & Drabman, R. (1971). Token reinforcement programs in the classroom: A review. Psychological Bulletin, 75, 379-398.

O'Leary, S.G., & Dubey, D.R. (1979). Applications of self-control procedures by children: A review. Journal of Applied Behaviour Analysis, 12, 449-466.

O'Leary, S.G., & O'Leary, K.D. (1976). Behaviour modification in the school. In H. Leitenberg (Ed.), Handbook of behaviour modification and behaviour therapy (pp. 475-515). New York: Prentice-Hall.

Ollendick, T.M., Matson, J.L., Esveldt-Dawson, K., & Shapiro, E.S. (1980). Increasing spelling achievement: An analysis of treatment proceedings using an alternating treatments design. Journal of Applied Behaviour Analysis, 13, 645-654.

Osterweis, R. (1968). Pictures as inspiration for creativity. English Journal, 57, 93-95.

Page, E.B. (1968). The use of the computer in analyzing student essays. International Review of Education, 14, 210-225.

Page, E.B., & Paulus, D. (1968). The analysis of essays by computer. Storrs, CT: University of Connecticut.

Paniagua, F.A., & Baer, D.M. (1982). The analysis of correspondence training as a chain reforceable at any point. Child Development, 53, 786-798.

Parkins, W.L. (1963). Motion pictures and written composition. English Journal, 52, 31-36.

Parnes, S.J. (1967). Creative behaviour guidebook. New York: Charles Scribner's Sons.

Parsonson, B.S., & Baer, D.M. (1978). Training generalized improvisation of tools by preschool children. Journal of Applied Behaviour Analysis, 11, 363-380.

Pearce, D.L. (1983). Guidelines for the use and evaluation of writing in content classrooms. Journal of Reading, 27, 212-218.

Perkins, K. (1980). Using objective methods of attained writing proficiency to discriminate among holistic evaluations. TESOL Quarterly, 14, 61-69.

Perkins, K. (1983). On the use of composition scoring techniques, objective measures and objective tests to evaluate ESL writing ability. TESOL Quarterly, 17, 651-671.

Perron, J.D. (1977, April). Written syntactic complexity and the modes of discourse. Paper presented at the Annual Meeting of the American Educational Research Association, New York.

Perron, J.D. (1979). Composition and cognition. English Education, 10, 144-154.

Personke, C. (1968). A gauntlet retrieved. Elementary English, 45, 464-467.

Petty, W.T. (1978). The writing of young children. In C.R. Cooper, & L. Odell (Eds), Research on composing: Points of departure (pp.73-83). Urbana, IL: National Council of Teachers of English.

Pinsent, P. (1984). Some current perspectives on the writing of young children. Early Child Development and Care, 14, 125-139.

Polloway, E.A., Patton, J.R., & Cohen, S.B. (1981). Written language for mildly handicapped students. Focus on Exceptional Children, 14(3), 1-16.

Poplin, M.S., Gray, R., Larsen, S., Banikowski, A., & Mehring, T. (1980). A comparison of components of written expression abilities in learning disabled and non-learning disabled students at three grade levels. Learning Disability Quarterly, 3, 46-53.

Poteet, J.A. (1979). Characteristics of written expression of learning disabled and non-learning disabled elementary school students. Diagnostique, 4, 60-74.

Potter, R.R. (1967). Sentence structure and prose quality: An exploratory study. Research in the Teaching of English, 1, 17-28.

Prater, D., & Padia, W. (1980). Effects of modes of discourse in writing performance in grades four and six. Paper presented at the Annual Meeting of the American Educational Research Association, Boston.

Prater, D.L., & Padia, W. (1983a). Developing parallel holistic and analytic scoring guides for assessing elementary writing samples. Journal of Research and Development in Education, 17, 20-24.

Prater, D., & Padia, W. (1983b). Effects of modes of discourse on writing performance in grades four and six. Research in the Teaching of English, 17, 127-134.

Quellmalz, E.S. (1984). Toward successful large-scale writing assessment: Where are we now? Where do we go from here? Educational Measurement: Issues and Practice, 3, 29-32.

Quellmalz, E.S., Capell, F.J., & Chou, C. (1982). Effects of discourse and response mode on the measurement of writing competence. Journal of Educational Measurement, 19, 241-258.

Rapport, M.D., & Bostow, D.E. (1976). The effects of access to special activities on performance in four categories of academic tasks with third grade students. Journal of Applied Behaviour Analysis, 9, 372.

Rayek, E., & Nesselroad, E. (1972). Application of behaviour principles to the teaching of writing, spelling and composition. In G. Semb (Ed.), Behaviour analysis and education (pp.171-184). The University of Kansas Support and Development Centre for Follow Through, Department of Human Development.

Raymond, J.C. (1982). What we don't know about the evaluation of writing. College Composition and Communication, 33, 399-403.

Read, C. (1981). Writing is not the inverse of reading for young children. In C.H. Frederiksen, & J.F. Dominic (Eds), Writing: The nature, development and teaching of written communication.(Vol. 2) Writing: Process, development and communication (pp.105-118). Hillsdale, NJ: Lawrence Erlbaum.

Read, W.H. (1969). Effects of mixed reinforcement contingencies on adults' control of children's behaviour. Journal of Applied Behaviour Analysis, 2, 249-253.

Reid, D.K., & Hresko, W.P. (1980). A developmental study of the relationship between oral language and early reading in learning disabled and normally achieving children. Learning Disability Quarterly, 3, 54-61.

Remondino, C. (1959). A factorial analysis of the evaluation of scholastic compositions in the mother tongue. British Journal of Educational Psychology, 29, 242-251.

Risley, T.R. (1977). The social context of self-control. In B.R. Stuart (Ed.), Behavioural self-management: Strategies, techniques and outcome (pp.71-81). New York: Brunner/Mazel.

Risley, T.R., & Hart, B. (1968). Developing correspondence between the nonverbal and verbal behaviour of preschool children. Journal of Applied Behaviour Analysis, 1, 267-281.

Rogers-Warren, A., & Baer, D.M. (1976). Correspondence between saying and doing: Teaching children to share and praise. Journal of Applied Behaviour Analysis, 9, 335-354.

Rogers-Warren, A., Warren, S.F., & Baer, D.M. (1977). A component analysis. Modeling, self-reporting, and reinforcement of self-reporting in the development of sharing. Behaviour Modification, 1, 307-322.

Rosen, H. (1973). Written language and the sense of audience. Educational Research, 15, 177-187.

Rosenbaum, M.E. (1968, November) An associative approach to creative composition. Paper presented to the NCTE sponsored conference on the Creative Process and Composition, Colorado Springs.

Rubin, R., & Buium, N. (1974, August). Language parameters in written compositions of 9-year-old children. Paper presented at the annual meeting of the American Psychological Association, New Orleans.

Russell, E.W. (1974). The power of behaviour control: A critique of behaviour modification methods. Journal of Clinical Psychology, 30, 111-136.

Sager, C. (1977). Improving the quality of written composition in the middle grades. Language Arts, 54, 760-762.

Salzberg, B.H., Wheeler, A.A., Devar, L.T., & Hopkins, B.C. (1971). The effect of intermittentfeedback and intermittent contingent access to play on printing of kindergarten children. Journal of Applied Behaviour Analysis, 4, 163-171.

San Jose, C.P.M. (1973). Grammatical structures in four modes of writing at the fourth grade level. Dissertation Abstracts International, 33, 5411A.

San Jose, C. (1978). Notes and comments. Research in the Teaching of English, 12, 91-92.

Sarbin, T.R. (1969). "On Zoellnerism". Notes on extending the talk-write metaphor. College English, 30, 645-667.

Scannell, D.P., & Marshall, J.C. (1966). The effect of selected composition errors on the grades assigned to essay examinations. American Educational Research Journal, 3, 125-130.

Schroeder, T.S. (1973). The effects of positive and corrective written teacher feedback on selected writing behaviours of fourth grade children. Dissertation Abstracts International, 34, 2935A.

Scriven, J.P., & Glynn, T. (1983). Performance feedback on written tasks for low-achieving secondary students. New Zealand Journal of Educational Studies, 18, 134-145.

Searle, D., & Dillon, D. (1980). The message of marking: Teacher written responses to student writing at intermediate grade levels. Research in the Teaching of English, 14, 233-242.

Sedlak, R.A., Cartwright, G.P. (1972). Written language abilities of EMR and nonretarded children with the same mental ages. American Journal of Mental Deficiency, 77, 95-99.

Seidman, E. (1968). Marking students' compositions: Implications of achievement motivation theory. Dissertation Abstracts International, 27, 2605A.

Shallcroft, D.J. (1981). Teaching creative behaviour: How to teach creativity in children of all ages. Englewood Cliffs, NJ: Prentice-Hall.

Shanahan, T. (1980). The impact of writing instruction on learning to read. Reading World, 19, 357-368.

Shanahan, T.(1984). Nature of the reading-writing relation: An exploratory multivariate analysis. Journal of Educational Psychology, 76, 466-477.

Sharpley, C.F., & Sharpley, A.M. (1981). Contingent versus noncontingent rewards in the classroom: A review of the literature. Journal of School Psychology, 19, 250-259.

Sharples, D. (1968). Content of creative writing. Elementary School Journal, 68, 419-426.

Shaughnessy, M.P. (1977). Errors and expectations. New York: Oxford University Press.

Sherman, J.A. (1964). Modification of nonverbal behaviour through reinforcement of related verbal behaviour. Child Development, 35, 717-723.

Sherwin, J.S. (1969). Four problems in teaching English: A critique of research. Scraton, PA: International Textbook.

Siegelman, L.A. (1983). The evaluation of written products in school age students: Techniques and developmental milestones. Dissertation Abstracts International, 43, 2315A.

Simms, R.B. (1983). Feedback - A key to effective writing. Academic Therapy, 19, 31-36.

Singh, N.N., Singh, J., & Winton, A.S.W. (1984). Positive practice overcorrection of oral reading errors. Behaviour Modification, 8, 23-37.

Singh, N.N., Winton, A.S.W., & Singh, J. (1985). Effects of delayed versus intermediate attention to oral reading errors on the reading proficiency of mentally retarded children. Applied Research in Mental Retardation, 6, 283-293.

Skinner, B.F. (1957). Verbal behaviour. New York: Appleton-Century-Crofts.

Slotnick, H.B. (1972). Toward a theory of computer essay grading. Journal of Educational Measurement, 9, 253-262.

Slotnick, H.B., & Knapp, J.V. (1971). Essay grading by computer: A laboratory phenomenon? English Journal, 60, 75-87.

Smith, F. (1975). The relation between spoken and written language. In E.H. Lenneberg, & E. Lenneberg (Eds), Foundations of language development: A multidisciplinary approach (Vol. 2) (pp.347-360). New York: Academic Press.

Smith, F. (1981). Myths of writing. Language Arts, 58, 792-798.

Smith, F. (1982). Writing and the writer. New York: Holt, Rinehart and Winston.

Smith, P.T. (1983). Patterns of writing errors in the framework of an information-processing model of writing. In D.R. Rogers, & J.A. Sloboda (Eds), Acquisition of symbolic skills. New York: Plenum Press.

Smith, V.H. (1969). Measuring teacher judgement in the evaluation of written composition. Research in the Teaching of English, 3, 181-195.

Smith, W.L. (1974). Syntactic recording of passages written at three levels of complexity. Journal of Experimental Education, 43, 66-72.

Sofell, C. (1929). A comparison of the use of imposed with self-chosen subjects in a creative writing program. Unpublished master's thesis, University of Pittsburgh.

Soloff, S. (1973). Effect of non-content factors on the grading of essays. Graduate Research in Education and Related Disciplines, 6, 44-54.

Solomon, D., & Rosenberg, L. (1964). Teacher-student feedback and classroom social structure. Journal of Social Psychology, 62, 197-210.

Sommers, N.I. (1979). The need for theory in composition research. College Composition and Communication, 30, 46-49.

Spaulding, R. (1963). Achievement, creativity and self-concept correlates of teacher-pupil transaction in elementary schools. Urbana, IL: University of Illinois Press.

Stahl, A. (1974). Structural analysis of children's compositions. Research in the Teaching of English, 8, 184-205.

Stallard, C. (1976). Composing: A cognitive process theory. College Composition and Communication, 27, 181-184.

Staudacher, C. (1968). Creative writing in the classroom. Palo Alto, CA: Fearon.

Stein, N.L. (1983). Methodological and conceptual issues in writing research. Elementary School Journal, 84, 100-108.

Steinberg, E.R. (1980). A garden of opportunities and a thicket of dangers. In L.W. Gregg, & E.R. Steinberg, (Eds), Cognitive processes in writing (pp.155-167). Hillsdale, NJ: Lawrence Erlbaum Associates.

Stevens, A.E. (1973). The effects of positive and negative evaluation on the written composition of low performing high school students. Dissertation Abstracts International, 34, 1178A.

Stewart, L.G.G. (1975). The differential effects of letter grades and written teacher comments on student objective test performance. Dissertation Abstracts International, 35, 6520-6521A.

Stewart, M.F. (1978). Syntactic maturity from high school to university: A first look. Research in the Teaching of English, 12, 37-46.

Stewart, M.F., & Grobe, C.H. (1979). Syntactic maturity, mechanics of writing, and teachers' quality ratings. Research in the Teaching of English, 13, 207-215.

Stiggins, R.J. (1982). A comparison of direct and indirect writing assessment methods. Research in the Teaching of English, 16, 101-114.

Stiggins, R.J., & Bridgeford, N.J. (1983). An analysis of published tests of writing proficiency. Educational Measurement: Issues and Practice, 2, 6-19.

Stokes, T.F., & Baer, D.M. (1977). An implicit technology of generalization. Journal of Applied Behaviour Analysis, 10, 349-367.

Strickland, R.G. (1960). Evaluating children's composition. Elementary English, 37, 321-330.

Stromer, R.S. (1977). Remediating academic deficiencies in learning disabled children. Exceptional Children, 43, 432-440.

Sweet, R.C. (1966). Educational attainment and attitudes toward school as a function of feedback in the form of teachers' written comments. Madison, WI, University of Wisconsin.

Taylor, W.F., & Hoedt, K.C. (1966). The effect of praise upon the quality and quantity of creative writing. Journal of Educational Research, 60, 80-83.

Tompkins, G.E. (1982). Seven reasons why children should write stories. Language Arts, 59, 718-721.

Torrance, E.P. (1965). Rewarding creative behaviour, experiments in classroom creativity. Englewood Cliffs, NJ: Prentice-Hall.

Torrance, E.P. (1967). Scientific views of creativity and factors affecting its growth. In J. Kagan (Ed.), Creativity and learning (pp.73-91). Boston: Houghton Mifflin.

Tracey, D.A., Briddell, D.W., & Wilson, G.T. (1974). Generalization of verbal conditioning to verbal and non-verbal behaviour: Group therapy with chronic patients. Journal of Applied Behaviour Analysis, 7, 391-402.

Traxler, A.E., & Anderson, H.A. (1935). The reliability of an essay test in English. School Review, 43, 534-539.

Treiber, F.A., & Lahey, B.B. (1983). Toward a behavioural model of academic remediation with learning disabled children. Journal of Learning Disabilities, 16, 111-116.

Trenberth, A.S., & Parsonson, B.S. (1985, August). Training and generalization of form diversity by primary school children. Paper presented at the First Joint NZP$_S$S-APS Conference, Christchurch, New Zealand.

Tripp, J.G. (1978). The positive approach: Response - evaluation of children's writing. Language Arts, 55, 358-361.

Van Houten, R. (1979). The performance feedback system: Generalization of effects across time. Child Behaviour Therapy, 1, 219-236.

Van Houten, R., Hill, S., & Parsons, M. (1975). An analysis of a performance feedback system: The effects of timing and feedback, public posting, and praise upon academic performance and peer interaction. Journal of Applied Behaviour Analysis, 8, 449-457.

Van Houten, R., & McKillop, C. (1977). An extension of the effects of the performance feedback system with secondary school students. Psychology in the Schools, 14, 480-484.

Van Houten, R., Morrison, E., Jarvis, R., & McDonald, M. (1974). The effects of explicit timing and feedback on compositional response rate in elementary school children. Journal of Applied Behaviour Analysis, 7, 547-555.

Van Houten, R., & Thompson, C. (1976). The effects of explicit timing on math performance. Journal of Applied Behaviour Analysis, 9, 227-230.

Van Houten, R., & Van Houten, J. (1977). The performance feedback system in the special education classroom: An analysis of public posting and peer comments. Behaviour Therapy, 8, 366-376.

Vargas, J.S. (1978). A behavioural approach to the teaching of composition. Behaviour Analyst, 1, 16-24.

Veal, L.R. (1974). Syntactic measures and rated quality in the writing of young children. Studies in Language Education (Report No. 8). Athens: University of Georgia.

Veal, L.R., & Hudson, S.A. (1983). Direct and indirect measures for lage-scale evaluation of writing. Research in the Teaching of English, 17, 290-296.

Veal, L.R., & Tillman, M. (1971). Mode of discourse variation in the evaluation of children's writing. Research in the Teaching of English, 5, 37-45.

Vygotsky, L.S. (1962). Thought and language. New York: Wiley.

Walker, H.M., & Buckley, N.K. (1972). Effects of reinforcement, punishment, and feedback upon academic response rate. Psychology in the Schools, 9, 186-193.

Walmsley, S.A. (1983). Writing disability. In P. Mosenthal, S. Walmsley, & L. Tamor (Eds), Research on writing: Principles and methods (pp.267-286). London: Longman.

Walmsley, S.A. (1984). Helping the learning disabled child overcome writing disabilities in the classroom. Topics in Learning and Learning Disabilities, 3, 81-89.

Watson, C. (1980). The effects of maturity and discourse type on the written syntax of superior high school seniors and upper level college English majors. Dissertation Abstracts International, 41, 141A.

Whitman, T.L., Scibak, J.W., Butler, K.M., Richter, R., & Johnson, M.R. (1982). Improving classroom behaviour in mentally retarded children through correspondence training. Journal of Applied Behaviour Analysis, 15, 545-564.

Wildman, R.W., & Wildman, R.W. (1975). Generalization of behaviour modification procedures: A review with special emphasis on classroom applications. Psychology in the Schools, 12, 432-438.

Wilkinson, A.M. (1964). Research on formal grammar. N.A.T.E. Bulletin, I, 24-26.

Willems, E.P. (1974). Behavioural technology and behavioural ecology. Journal of Applied Behaviour Analysis, 7, 151-165.

Williams, J.A., & Stokes, T.F. (1982). Some parameters of correspondence training and generalized verbal control. Child and Family Behaviour Therapy, 4, 11-32.

Williams, J.M. (1980). Non-linguistic linguistics and the teaching of style. Language and Style, 13, 24-40.

Wilson, M.G. (1976). Self-management of story-writing in a special class: Fading of teacher models and shaping independent writing. Master's thesis, Department of Education, University of Auckland.

Winett, R.A., & Roach, A.M. (1973). The effects of reinforcing academic performance on social behaviour: A brief report. Psychological Record, 23, 391-396.

Winett, R.A., & Winkler, R.C. (1972). Current behaviour modification in the classroom: Be still, be quiet, be docile. Journal of Applied Behaviour Analysis, 5, 499-504.

Wing, A.M., & Baddeley, A.D. (1980). Spelling errors in handwriting: A corpus and a distributional analysis. In U. Frith (Ed.), Cognitive processes in spelling (pp.251-285). London: Academic Press.

Wingfield, R.J. (1975). Five ways of dealing with errors in written composition. English Language Teaching Journal, 29, 311-313.

Wiseman, D.L. (1984). Helping children take early steps toward reading and writing. Reading Teacher, 37, 340-344.

Witte, S.P. (1980). Toward a model for research in written composition. Research in the Teaching of English, 14, 73-81.

Witte, S.P., & Davis, A.S. (1980). The stability of T-unit length: A preliminary investigation. Research in the Teaching of English, 14, 5-17.

Witty, P., Martin, W. (1957). An analysis of children's compositions written in response to a film. Elementary English, 34, 158-163.

Wotherspoon, A.T. (1974). Modification of writing behaviour in a special class. Unpublished master's thesis, Department of Education, University of Auckland.

Wyatt, N.M. (1961). A study of the relationship of extensive reading to certain writing skills of a selected group of sixth-grade children. University of Kansas Bulletin of Education, 16, 13-18.

Young, R.E. (1978). Paradigms and problems: Needed research in rhetorical invention. In C.R. Cooper, & L. Odell (Eds), Research on composing: Points of departure (pp.29-47). Urbana, IL: National Council of Teachers of English.

Young, R. (1981). Problems and the composing process. In C.H. Frederiksen, & J.F. Dominic (Eds), Writing: The nature, development and teaching of written communication. (Vol. 2) Writing: Process development, and communication (pp.59-66). Hillsdale, NJ: Lawrence Earlbaum.

Zoellner, R. (1969a). Talk-write: A behavioural pedagogy for composition. College English, 30, 267-320.

Zoellner, R. (1969b). Response: Mentalizing S.R. Rodent. College English, 31, 215-230.

APPENDIX A

Picture Prompt Topics for Experiment 2

Sessions	Picture Prompt Type	
	Theme	Action
1. Baseline		
1.	House	
2.		Men building on construction site.
3.		Children playing in river.
4.		Children cooking.
5.	Dog	
6.		Runners.
7.	Chair	
8.	Child in bed	
9.	Tray of eggs	
10.		Unloading bags off airport bus.
2. Intervention I		
11.		Diving scene at lake.
12.		Farmer and dog rounding up sheep.
13.	Father and son on step	
14.	Helicopter	
15.	Alarm clock	
16.		Boys playing rugby.
3. Intervention II		
17.		Playing with balloons at party.
18.	Leopard	
19.		Fishing boat activities.
20.	Modern car	
21.		Three different sports players.
22.	Lady in fancy clothes	
23.		Child jumping over barrel on horse
24.	Set of cards	
4. Intervention III		
25.		Dog chasing boy on bike.
26.		Old man tripping on hose.
27.	Family camping	
28.	Two clowns	
29.	Family breakfast	
30.		Native spearing a fish.
31.		Indian stalking buffalo
32.	Classroom of children	

APPENDIX B

Poster Topics for Experiment 1

Session	Topic	Session	Topic
1. Baseline			
1.	motorbike	42.	travel poster
2.	tribesmen	43.	old police type car
3.	mechanic	44.	motorbike promotion
4.	skateboarding	45.	cruise liner
5.	BMX rider	46.	aeroplane in flight
6.	airplane	47.	city bus in Vienna
7.	old house	48.	fire in boy's bedroom
8.	grader	49.	explorers
9.	BMX riding		
10.	fire engine		
11.	cathedral	3. Transfer . . . Maintenance	
		50.	cable car in Italy
2. Intervention		51.	forestry poster
12.	Fiji	52.	train
13.	Queenstown	53.	transport poster
14.	island (deserted)	54.	Australian travel poster
15.	highway	55.	horse
16.	milk truck	56.	girl in bikini
17.	pointsman	57.	Big M poster
18.	road repairers	58.	livesaver on beach
19.	Mt. Cook	59.	operating theatre
20.	Heliskiing	60.	sailing scene
21.	snow mobile	61.	airport
22.	drinking scene	62.	skiing
23.	crash scene	63.	gliding
24.	car	64.	"Come a Hot Friday"
25.	accident	65.	beach scene
26.	anti-smoking	66.	sea coast
27.	electricity	67.	forest parks
28.	ship	68.	Papua New Guinea
29	Shotover River	69.	war theme
30.	boy and ball	70.	railway station
31.	motorbike safety		
32.	dam		
33.	car accident		
34.	fashion		
35.	food		
36.	farm in country		
37.	ghost		
38.	drinking while sailing		
39.	bus tour-coach		
40.	skiing Coronet Peak		
41.	hang gliding		